The Student Equation: Unpacking the factors that add up to success

IE 6200 Term Project

Professor Rajesh Jugulum

College of Engineering, Northeastern University

Aditi Joshi, Jayendra Deshmukh, Layashree Adepu, Tarang Garg

December 04, 2023

Contents

• Abstract

This data analysis delves into the complex interplay between various factors and their influence on students' academic performance. The study categorizes these factors into four dimensions: social, environmental, physiological, and psychological. Variables such as anxiety level, self-esteem, mental health history, depression, headache, blood pressure, sleep quality, breathing problems, noise level, living conditions, safety, basic needs, study load, teacher-student relationship, future career concerns, social support, peer pressure, extracurricular activities, and bullying are scrutinized to discern their collective impact on academic outcomes.

Using advanced statistical methods, including correlation analyses and regression models, we aim to unveil the nuanced ways in which these factors positively or negatively affect student academic performance. The research integrates these variables within the broader context of social, environmental, physiological, and psychological dimensions, providing a holistic understanding of the intricate dynamics at play.

For instance, we explore whether social support and extracurricular activities positively correlate with academic success within the social dimension, or if physiological factors such as sleep quality and blood pressure impact academic performance. This study not only considers individual variables but also investigates their collective influence within each dimension.

The findings of this research are that academic performance hinges on certain key factors, like anxiety level, bullying, quality of sleep. Ultimately, our goal is to contribute to the development of targeted interventions and support systems that enhance overall student well-being and academic success.

• Methodology

The system under investigation involves a comprehensive examination of various factors affecting students, organized into four distinct dimensions: social, environmental, physiological, and psychological. This system encompasses a multitude of variables, each representing a crucial aspect of the student's experience.

A) Data collected
We have collected the data from Kaggle.

Student Stress Factors: A Comprehensive Analysis

This dataset contains around 20 features that create the most impact on the Stress of a Student. The features are selected scientifically considering 5 major factors, they are Psychological, Physiological, Social, Environmental, and Academic Factors. Some of them are:
Psychological Factors => 'anxiety_level', 'self_esteem', 'mental_health_history', 'depression'

Physiological Factors => 'headache', 'blood_pressure', 'sleep_quality', 'breathing_problem.
Environmental Factors => 'noise_level', 'living_conditions', 'safety', 'basic_needs'.
Academic Factors => 'academic_performance'.
Social Factor => 'social_support', 'peer_pressure', 'extracurricular_activities', 'bullying'

This analysis focuses on a subset of five available parameters, deliberately selecting four for in-depth exploration. Within the academic domain, our primary focus rests upon academic performance, which will serve as the dependent variable on the Y axis. This strategic selection allows for a more nuanced and focused investigation of the relationships between the chosen parameters and academic achievement.

B) System Explained

The data collected here is focused on various parameters that contribute to the academic performance of students. We took this dataset from Kaggle that was computed from January 2022 to October 2022. We are looking at the anxiety levels that

Components of the System:

It refers to the fundamental elements that constitute the intricate framework under examination.

1. **Social Dimension Components:**
    a. Variables such as social support, peer pressure, and extracurricular activities are key elements within the social dimension. These factors reflect the interpersonal and communal aspects of students' lives that may influence their academic performance.
2. **Environmental Dimension Components:**
    a. Living conditions, safety, and noise level constitute essential components within the environmental dimension. These factors highlight the external surroundings and contextual factors that students navigate in their academic environments.
3. **Physiological Dimension Components:**
    a. Anxiety level, blood pressure, sleep quality, and breathing problems represent crucial components within the physiological dimension. These variables delve into the physical and health-related aspects that may impact students academically.
4. **Psychological Dimension Components:**
    a. Self-esteem, mental health history, depression, and future career concerns are integral components within the psychological dimension. These variables shed light on the intricate landscape of students' thoughts, emotions, and aspirations in the academic realm.

Random Variables:

These variables are deemed as sources of randomness, and their values are not fixed but can vary across individuals or situations within the student population. In the broader scope of statistical analysis, random variables provide a way to model and quantify uncertainty.

Elaborating on the concept of random variables within the study's framework:

- **Anxiety Level:**
  - Anxiety level is a random variable as it can vary among students, and its exact value is uncertain. Some students may experience high anxiety, while others may have lower levels.
- **Sleep Quality:**
  - Sleep quality is a random variable because it varies across individuals. Some students may consistently experience good sleep quality, while others may face fluctuations.
- **Social Support:**
  - The level of social support is a random variable as it can differ among students. Some may have strong social support networks, while others may have limited or variable levels of support.
- **Study Load:**
  - Study load is a random variable as it can differ from one student to another. Students may have varying academic workloads based on their courses, extracurricular activities, and personal commitments.
- **Safety Perception:**
  - How safe students perceive their surroundings is a random variable. This perception can be influenced by various factors and may vary from student to student.

Assumptions:

These assumptions play a crucial role in the interpretation of results and the validity of statistical inferences.

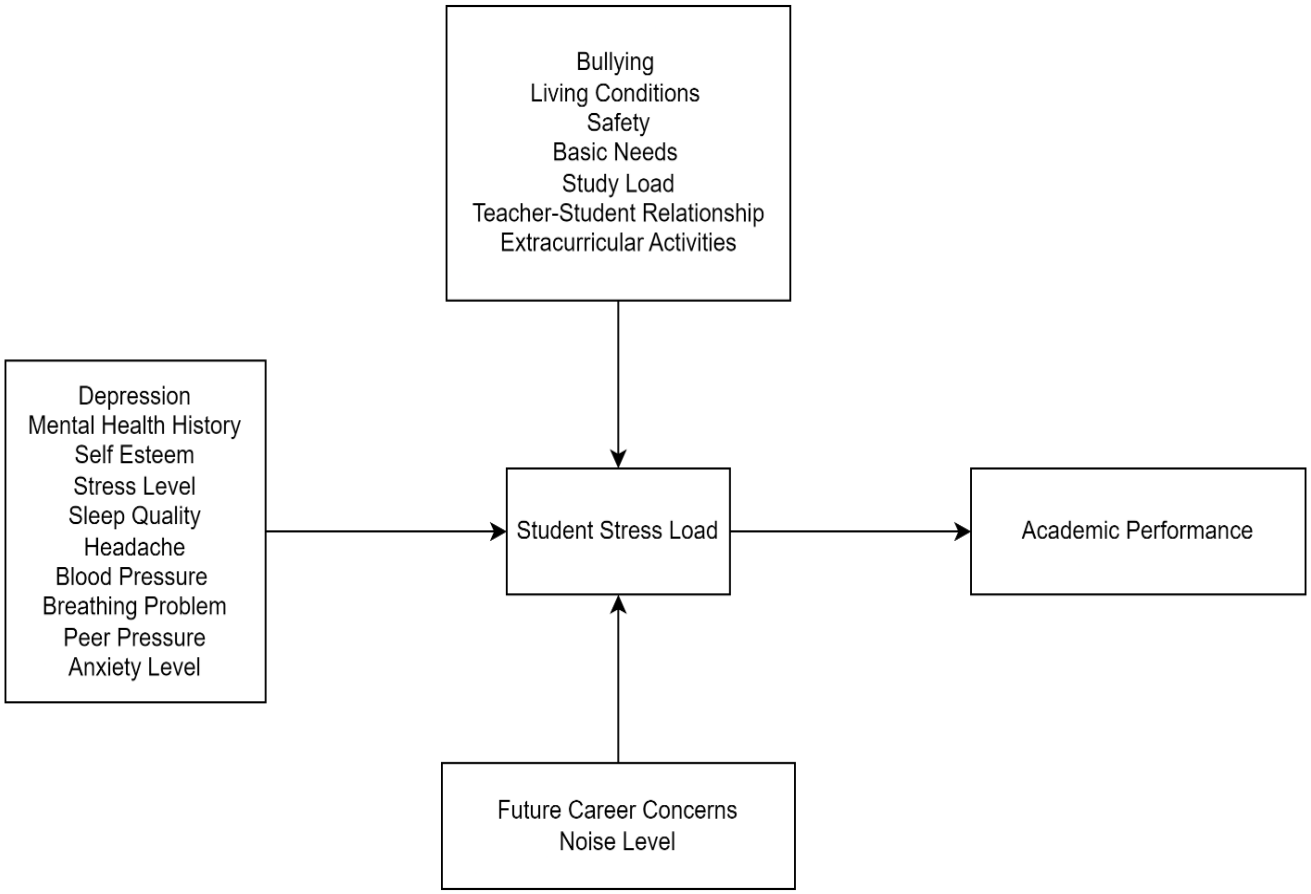
Here are some potential assumptions that might be relevant to provided data:

- **Normality of Data:**
  - An assumption could be that the data distribution of certain variables, such as anxiety levels or sleep quality, is approximately normal.
- **Independence of Observations:**
  - It might be assumed that each student's response or data point is independent of others.
- **Normality of Residuals:**
  - For regression analyses, it may be assumed that the residuals (the differences between observed and predicted values) are normally distributed.

Purpose of Analysis:

In summary, the described system provides a methodical and inclusive approach to studying the factors that shape students' academic experiences. By considering social, environmental, physiological, and psychological dimensions collectively, the system aims to contribute nuanced and valuable insights that go beyond isolated examinations of individual variables. This holistic perspective is essential for comprehensively understanding the various influences on students and improving the overall educational experience.

- **P-Diagram**



Inputs: Depression, Mental Health History, Self Esteem, Stress Level, Sleep Quality, Headache, Blood pressure, Breathing Problem, Peer Pressure, Anxiety Level

Control Factors: Bullying, Living Conditions, Safety, Basic Needs, Study Load, Teacher-Student Relationship, Extracurricular Activities

Noise Factors: Future Career Concerns, Noise Level

Output: Academic Performance

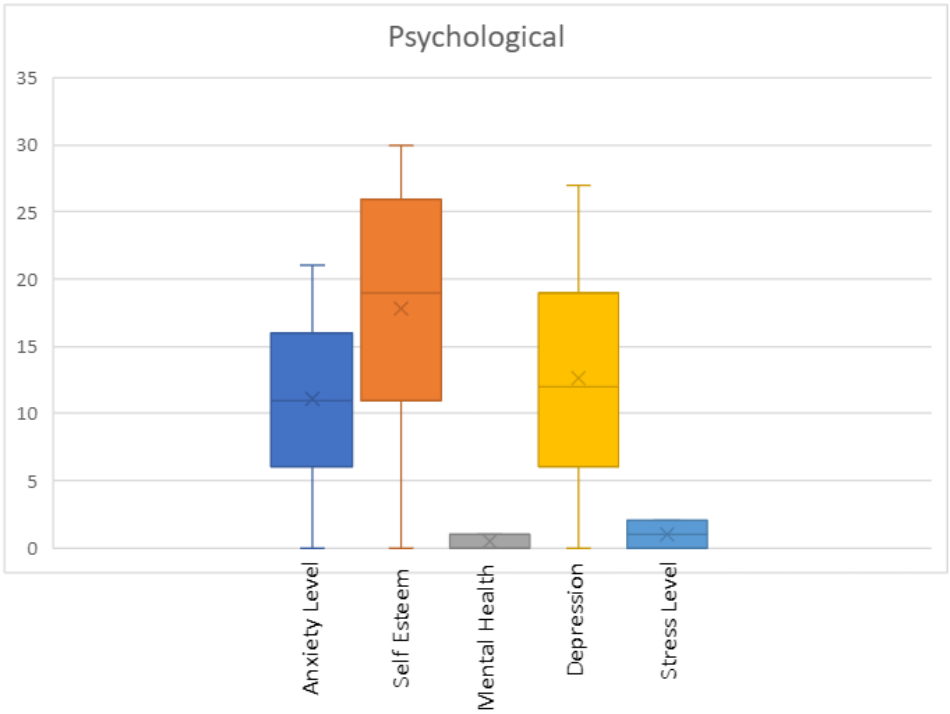
- **Data**

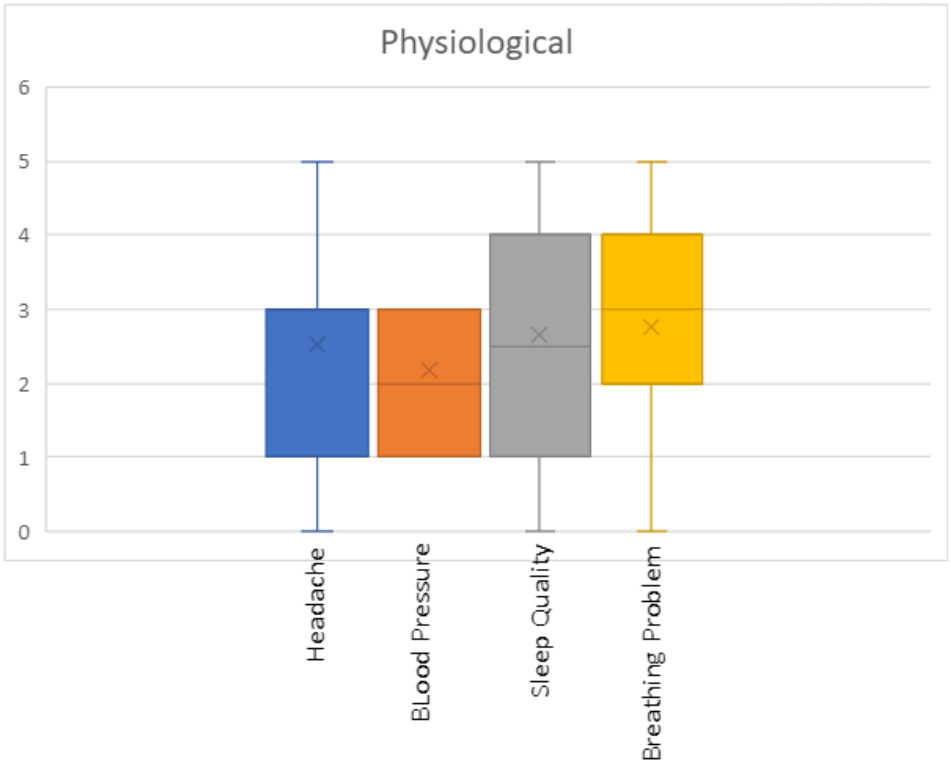
Exploring the Impact of Social, Environmental, Physiological, and Psychological Factors on Students' Academic Performance	Academics	academic_performance	3	1	2	2	4	2	5	1	3	2	1	1	5	1	4	3	5	4	2	1	5	2	4	1	
	Psychological	anxiety_level	14	15	12	16	16	20	4	17	13	6	17	17	5	9	2	11	6	7	11	21	3	18	7	20	
		self_esteem	20	8	18	12	28	13	26	3	22	8	12	15	28	23	28	21	28	25	23	1	27	1	27	5	
		mental_health_history	0	1	1	1	0	1	0	1	1	0	1	1	0	1	0	0	0	0	0	1	0	1	0	1	
		depression	11	15	14	15	7	21	6	22	12	27	25	22	8	24	3	14	1	3	12	25	0	21	5	26	
		stress_level	1	2	1	2	1	2	0	2	1	1	2	2	0	2	0	1	0	0	1	2	0	2	0	2	
	Physiological	headache	2	5	2	4	2	3	1	4	3	4	4	3	1	4	1	3	1	1	3	4	1	4	1	3	
		blood_pressure	1	3	1	3	3	3	2	3	1	3	3	3	2	3	2	1	2	2	1	3	2	3	2	3	
		sleep_quality	2	1	2	1	5	1	4	1	2	1	1	1	4	1	4	2	4	4	2	1	4	1	4	1	
		breathing_problem	4	4	2	3	1	4	1	5	4	2	3	5	2	0	2	4	2	2	2	4	1	3	1	4	
	Environmental	noise_level	2	3	2	4	3	3	1	3	3	0	4	5	2	1	1	2	1	2	3	4	1	5	1	4	
		living_conditions	3	1	2	2	2	2	4	1	3	5	2	2	3	2	3	2	4	4	2	1	3	1	3	2	
		safety	3	2	3	2	4	2	4	1	3	2	1	1	5	4	4	2	5	5	3	2	5	1	5	1	
		basic_needs	2	2	2	2	3	1	4	1	3	2	1	1	5	3	4	2	4	4	3	1	4	2	5	2	
	Social	social_support	2	1	2	1	1	1	3	1	3	1	1	1	3	0	3	2	3	3	3	1	3	1	3	1	
		peer_pressure	3	4	3	4	5	4	2	4	3	5	4	5	1	1	1	3	2	1	3	4	1	4	1	5	
		extracurricular_activities	3	5	2	4	0	4	2	4	2	3	4	5	1	0	2	2	2	1	2	4	2	4	2	4	

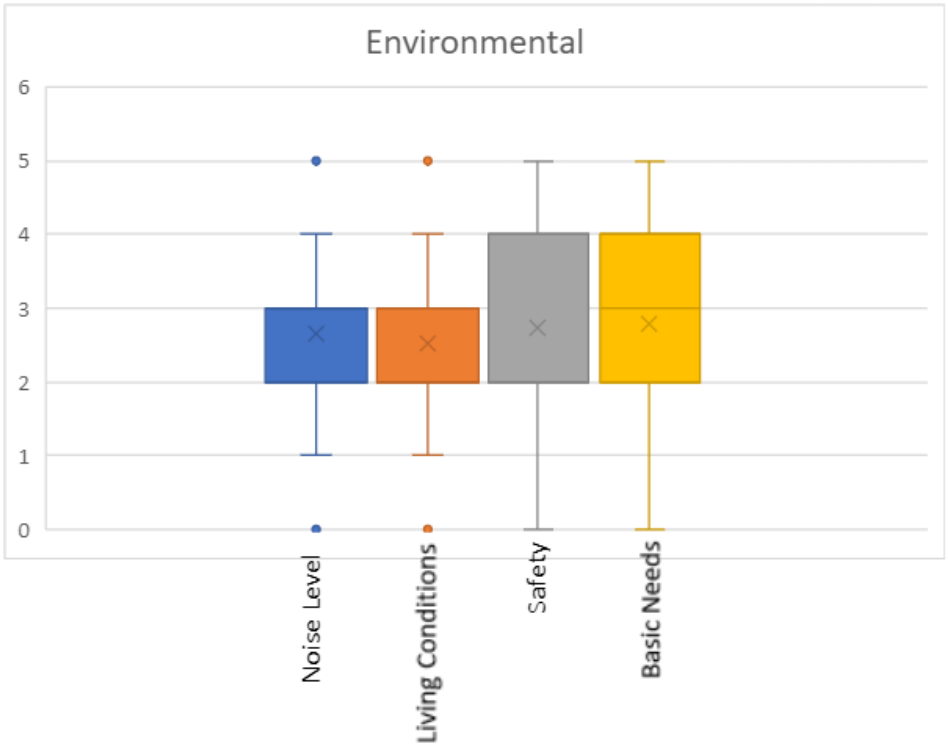
• Descriptive Statistics

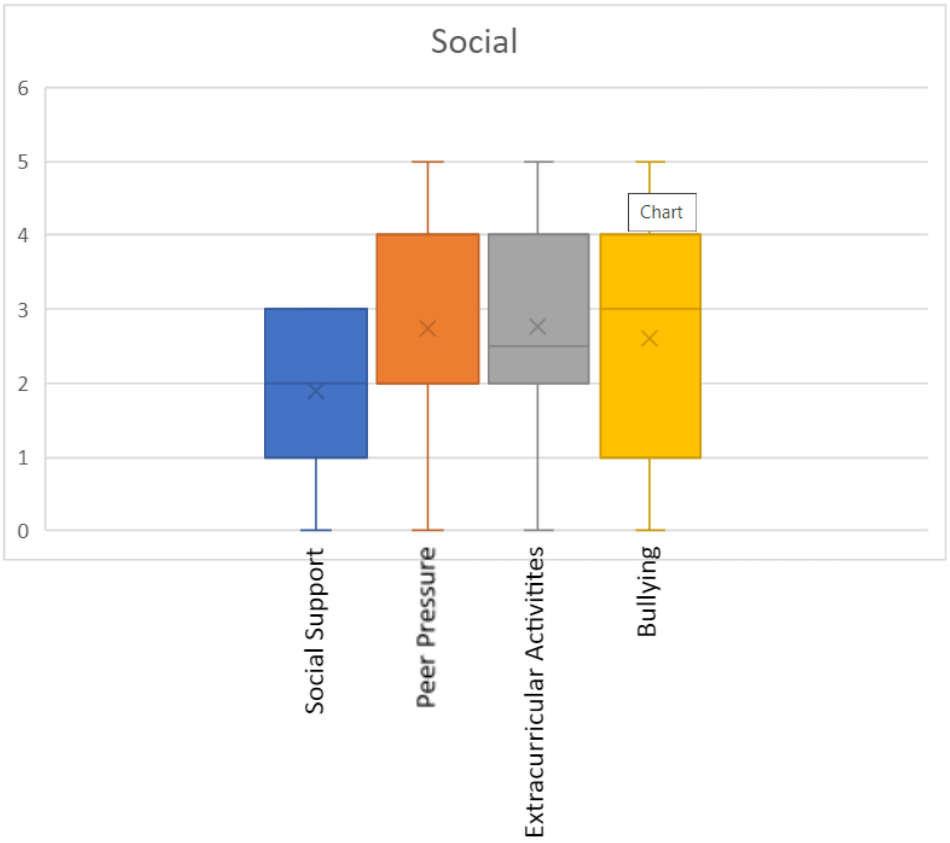
We conducted a thorough analysis for each individual factor by delving into descriptive statistics, unveiling crucial insights into the data distribution. This involved examining measures like the mean, variance, and standard deviation, offering a comprehensive view of central tendencies and the spread of values within each factor. Moreover, we explored additional statistical parameters such as the maximum and minimum values, as well as skewness, to capture the asymmetry of the data distribution. To visually represent these findings, we employed box plots, providing a clear visualization of the dataset's distribution and revealing any potential outliers. Notably, our scrutiny uncovered outliers in a limited number of factors, specifically in noise level, living conditions, and study load. These outliers were detected on both ends of the range, indicating extreme values beyond the typical data spread for these factors.

BOX PLOTS




Physiological
6
5
4
3
2
1
0
Headache
BLood Pressure
Sleep Quality
Breathing Problem

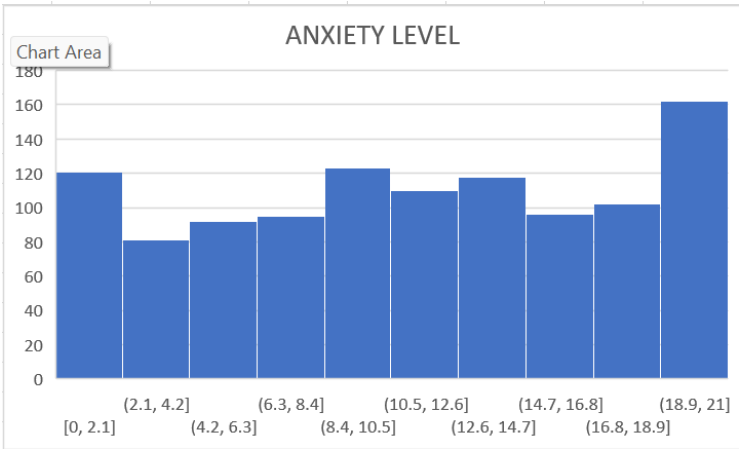

Environmental
6
5
4
3
2
1
0
Noise Level
Living Conditions
Safety
Basic Needs


Social
6
5
4
3
2
1
0
Chart
Social Support
Peer Pressure
Extracurricular Activitites
Bullying

HISTOGRAMS

Anxiety Level

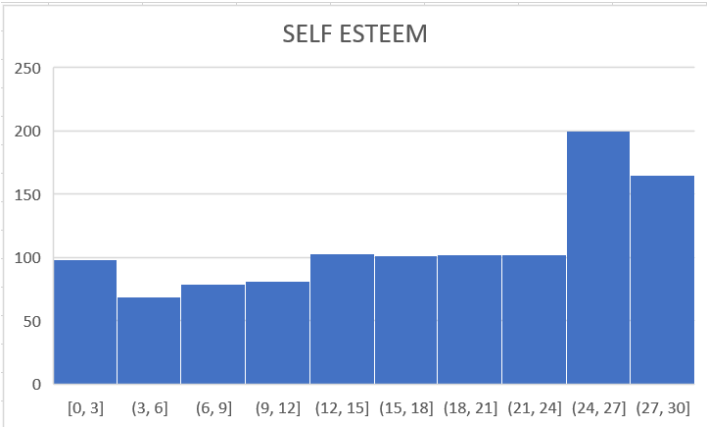
anxiety_level	
Mean	11.06364
Standard Error	0.184451
Median	11
Mode	13
Standard Deviation	6.117558
Sample Variance	37.42452
Kurtosis	-1.09413
Skewness	-0.08261
Range	21
Minimum	0
Maximum	21
Sum	12170
Count	1100



In summary, the dataset demonstrates a diverse range of anxiety levels, with a central tendency around 11.06. The distribution is moderately spread, and while there is a slight skew to the left, the overall shape is relatively flat.

Self Esteem

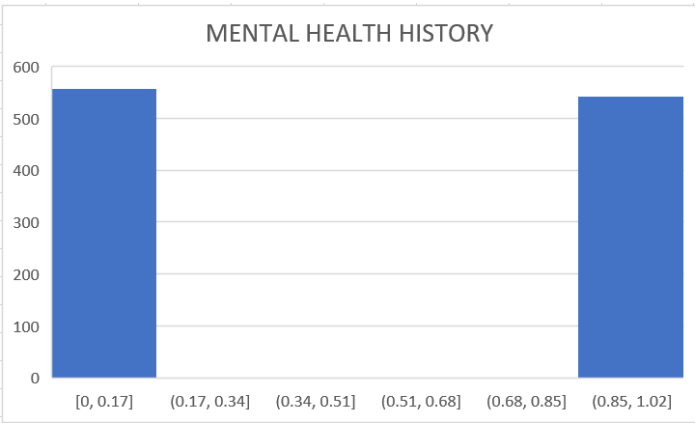
self_esteem	
Mean	17.77727
Standard Error	0.26969
Median	19
Mode	25
Standard Deviation	8.944599
Sample Variance	80.00585
Kurtosis	-1.06183
Skewness	-0.39654
Range	30
Minimum	0
Maximum	30
Sum	19555
Count	1100



In summary, the dataset for self-esteem reveals a diverse range of scores with a central tendency around 17.78. The distribution is moderately spread, and there is a slight leftward skew with a flatter peak. The mode at 25 suggests a common self-esteem score in the dataset.

Mental Health History

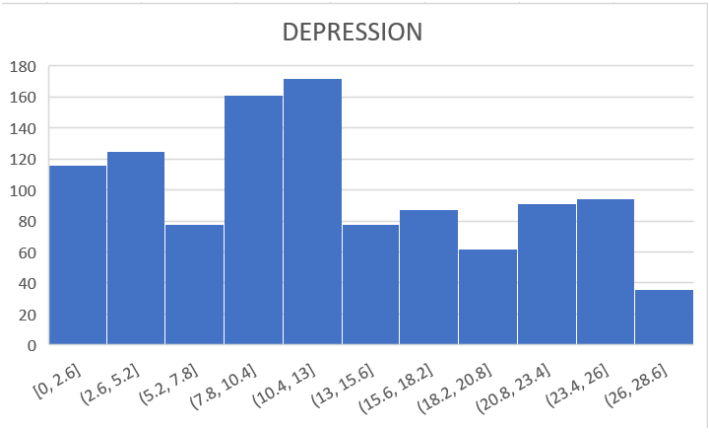
mental_health_history	
Mean	0.492727
Standard Error	0.015081
Median	0
Mode	0
Standard Deviation	0.500175
Sample Variance	0.250175
Kurtosis	-2.0028
Skewness	0.029134
Range	1
Minimum	0
Maximum	1
Sum	542
Count	1100



In summary, the dataset for mental health history indicates a predominantly binary distribution, with most individuals reporting no mental health history (mode at 0). The central tendency is around 0.49, and there is a slight rightward skew with a sharper peak.

Depression

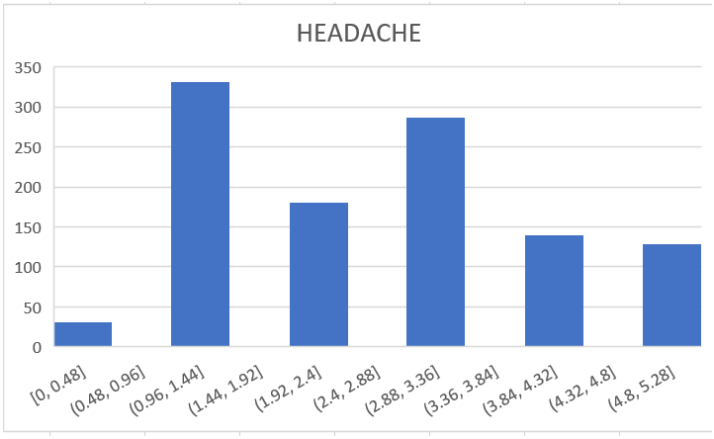
depression	
Mean	12.55545
Standard Error	0.232978
Median	12
Mode	10
Standard Deviation	7.727008
Sample Variance	59.70666
Kurtosis	-0.98884
Skewness	0.216638
Range	27
Minimum	0
Maximum	27
Sum	13811
Count	1100



In summary, the dataset for depression reveals a diverse range of scores with a central tendency around 12.56. The distribution is moderately spread, and there is a slight rightward skew with a flatter peak. The mode at 10 suggests a common depression level in the dataset.

Headache

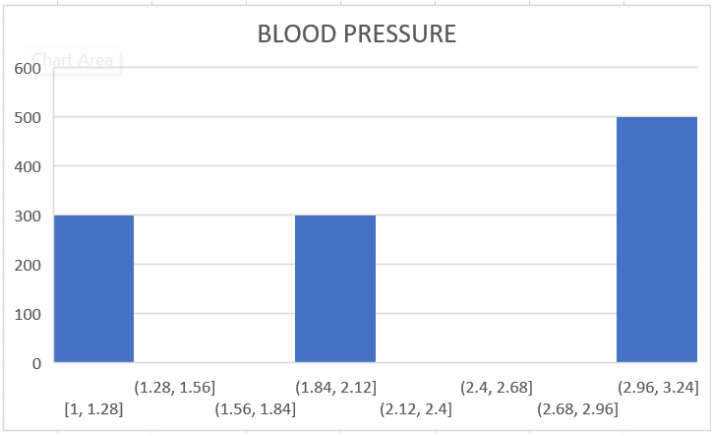
headache	
Mean	2.508182
Standard Error	0.042494
Median	3
Mode	1
Standard Deviation	1.409356
Sample Variance	1.986284
Kurtosis	-0.99659
Skewness	0.270494
Range	5
Minimum	0
Maximum	5
Sum	2759
Count	1100



In summary, the dataset for headaches reveals a diverse range of scores with a central tendency around 2.51. The distribution is moderately spread, and there is a slight rightward skew with a flatter peak.

Blood Pressure

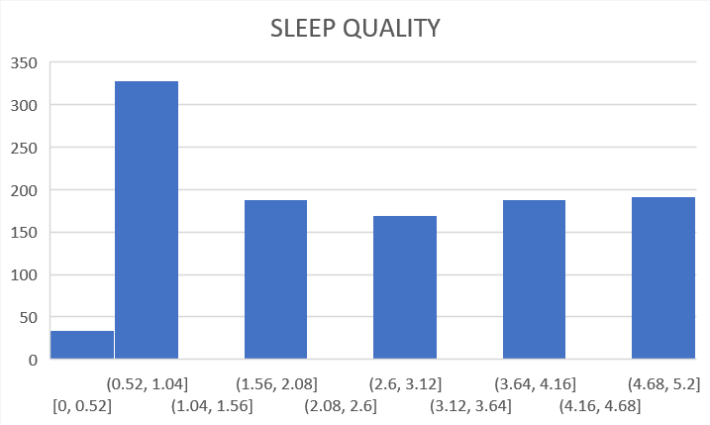
blood_pressure	
Mean	2.181818
Standard Error	0.025133
Median	2
Mode	3
Standard Deviation	0.833575
Sample Variance	0.694847
Kurtosis	-1.47403
Skewness	-0.35119
Range	2
Minimum	1
Maximum	3
Sum	2400
Count	1100



In summary, the dataset for blood pressure reveals a moderately narrow range of scores with a central tendency around 2.18. The distribution is moderately spread, and there is a slight leftward skew with a flatter peak. The mode at 3 suggests a common blood pressure level in the dataset.

Sleep Quality

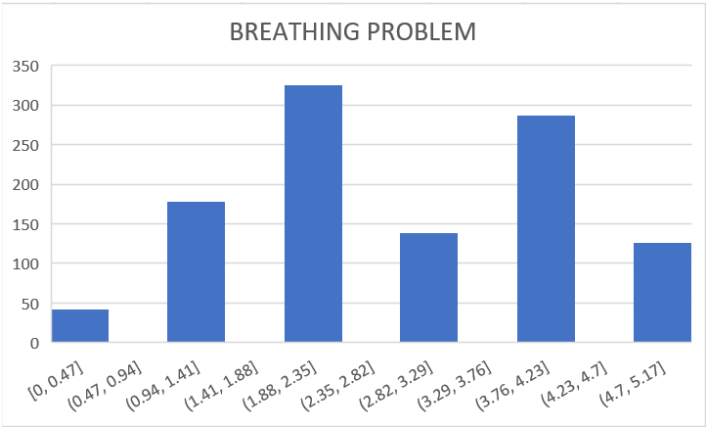
sleep_quality	
Mean	2.66
Standard Error	0.046685
Median	2.5
Mode	1
Standard Deviation	1.548383
Sample Variance	2.397489
Kurtosis	-1.32205
Skewness	0.177975
Range	5
Minimum	0
Maximum	5
Sum	2926
Count	1100



In summary, the dataset for sleep quality reveals a moderately wide range of scores with a central tendency around 2.66. The distribution is moderately spread, and there is a slight rightward skew with a flatter peak. The mode at 1 suggests a common sleep quality level in the dataset.

Breathing Problem

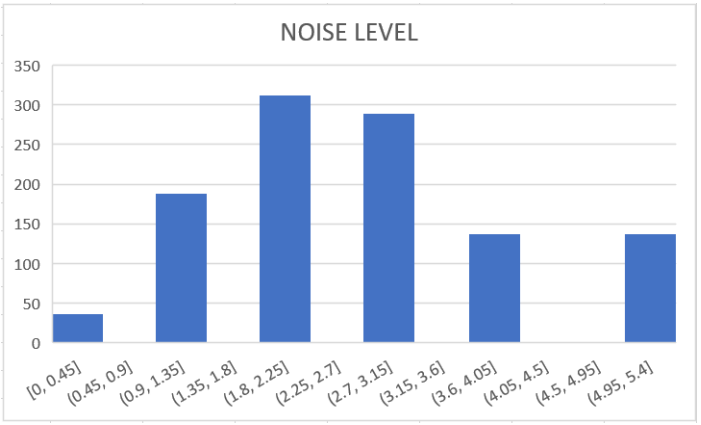
breathing_problem	
Mean	2.753636
Standard Error	0.042233
Median	3
Mode	2
Standard Deviation	1.400713
Sample Variance	1.961998
Kurtosis	-1.06004
Skewness	-0.00162
Range	5
Minimum	0
Maximum	5
Sum	3029
Count	1100



In summary, the dataset for breathing problems reveals a moderately wide range of scores with a central tendency around 2.75. The distribution is moderately spread, nearly symmetrical, and there is a mode at 2, suggesting a common level of breathing problems in the dataset.

Noise Level

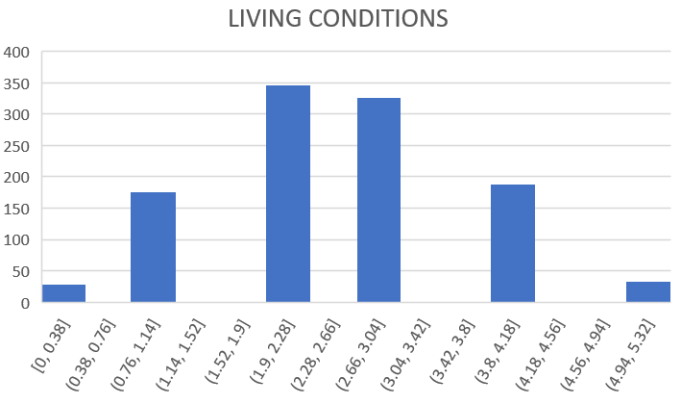
noise_level	
Mean	2.649091
Standard Error	0.040045
Median	3
Mode	2
Standard Deviation	1.328127
Sample Variance	1.763921
Kurtosis	-0.69576
Skewness	0.207122
Range	5
Minimum	0
Maximum	5
Sum	2914
Count	1100



In summary, the dataset for noise level reveals a moderate range of scores with a central tendency around 2.649. The distribution is moderately spread, and there is a slight rightward skew with a flatter peak. The mode at 2 suggests a common noise level score in the dataset.

Living Conditions

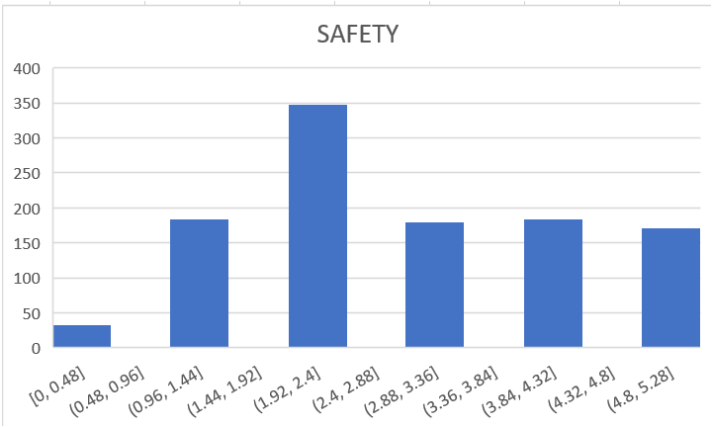
living_conditions	
Mean	2.518182
Standard Error	0.033745
Median	2
Mode	2
Standard Deviation	1.119208
Sample Variance	1.252626
Kurtosis	-0.49692
Skewness	0.018794
Range	5
Minimum	0
Maximum	5
Sum	2770
Count	1100



In summary, the dataset for living conditions reveals a moderate range of scores with a central tendency around 2.518. The distribution is moderately spread, and there is a very slight rightward skew with a flatter peak. The mode at 2 suggests a common living conditions score in the dataset.

Safety

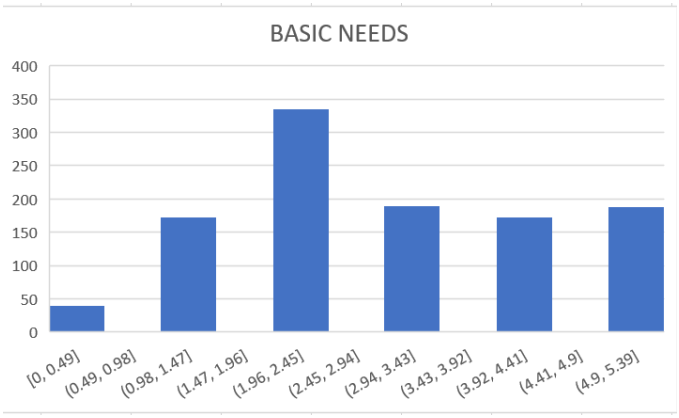
safety	
Mean	2.737273
Standard Error	0.042398
Median	2
Mode	2
Standard Deviation	1.406171
Sample Variance	1.977317
Kurtosis	-0.99983
Skewness	0.188097
Range	5
Minimum	0
Maximum	5
Sum	3011
Count	1100



In summary, the dataset for safety reveals a moderate range of scores with a central tendency around 2.737. The distribution is moderately spread, and there is a slight rightward skew with a flatter peak. The mode at 2 suggests a common safety score in the dataset.

Basic Needs

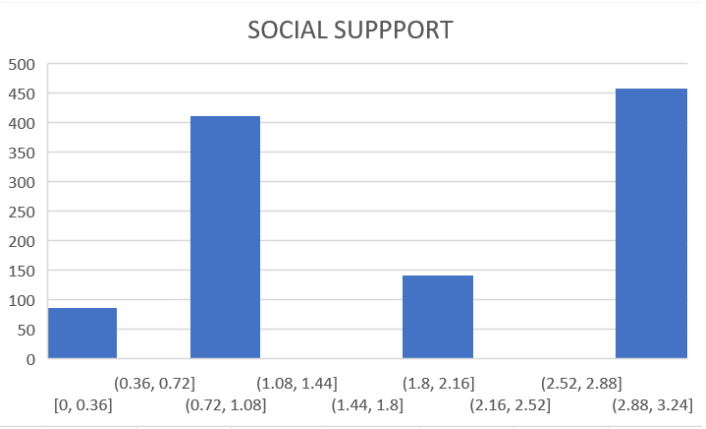
basic_needs	
Mean	2.772727
Standard Error	0.04323
Median	3
Mode	2
Standard Deviation	1.433761
Sample Variance	2.05567
Kurtosis	-1.00852
Skewness	0.135859
Range	5
Minimum	0
Maximum	5
Sum	3050
Count	1100



In summary, the dataset for basic needs reveals a moderate range of scores with a central tendency around 2.773. The distribution is moderately spread, and there is a slight rightward skew with a flatter peak. The mode at 2 suggests a common basic needs score in the dataset.

Social Support

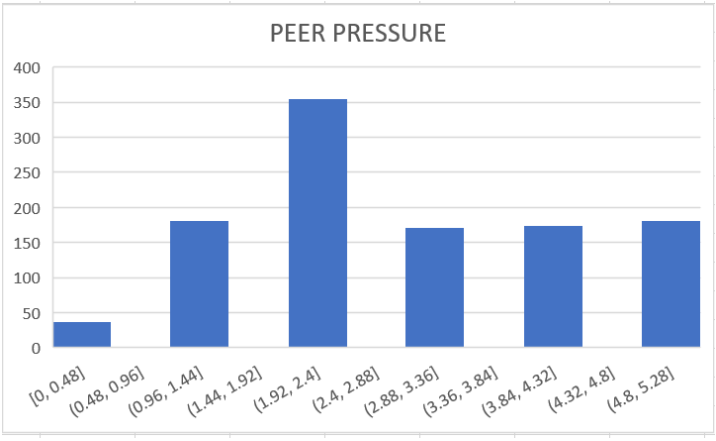
social_support	
Mean	1.881818
Standard Error	0.031593
Median	2
Mode	3
Standard Deviation	1.047826
Sample Variance	1.09794
Kurtosis	-1.43815
Skewness	-0.18096
Range	3
Minimum	0
Maximum	3
Sum	2070
Count	1100



In summary, the dataset for social support reveals a moderate range of scores with a central tendency around 1.882. The distribution is moderately spread, and there is a slight leftward skew with a flatter peak. The mode at 3 suggests a common social support score in the dataset.

Peer Pressure

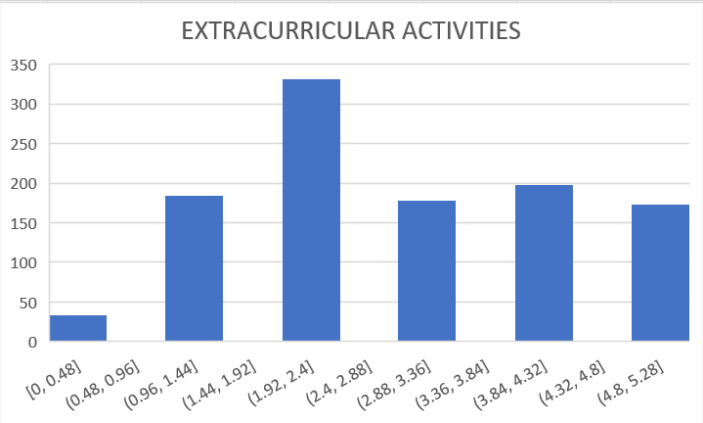
peer_pressure	
Mean	2.734545
Standard Error	0.042973
Median	2
Mode	2
Standard Deviation	1.425265
Sample Variance	2.031381
Kurtosis	-1.00718
Skewness	0.194719
Range	5
Minimum	0
Maximum	5
Sum	3008
Count	1100



In summary, the dataset for peer pressure reveals a moderate range of scores with a central tendency around 2.735. The distribution is moderately spread, and there is a slight rightward skew with a flatter peak. The mode at 2 suggests a common peer pressure score in the dataset.

Extracurricular Activities

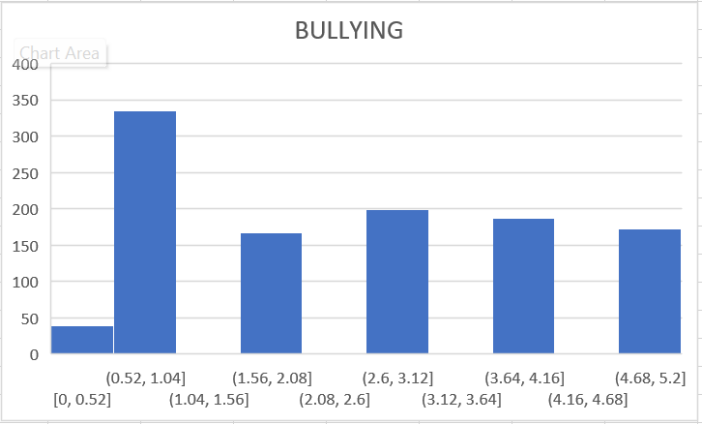
extracurricular_activities	
Mean	2.767273
Standard Error	0.042741
Median	2.5
Mode	2
Standard Deviation	1.417562
Sample Variance	2.009483
Kurtosis	-1.04031
Skewness	0.135604
Range	5
Minimum	0
Maximum	5
Sum	3044
Count	1100



In summary, the dataset for extracurricular activities reveals a moderate range of scores with a central tendency around 2.767. The distribution is moderately spread, and there is a slight rightward skew with a flatter peak. The mode at 2 suggests common extracurricular activities score in the dataset.

Bullying

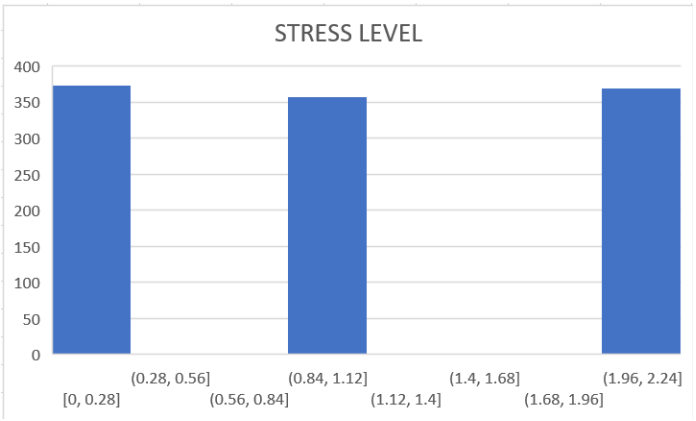
bullying	
Mean	2.617273
Standard Error	0.04616
Median	3
Mode	1
Standard Deviation	1.530958
Sample Variance	2.343832
Kurtosis	-1.27448
Skewness	0.175348
Range	5
Minimum	0
Maximum	5
Sum	2879
Count	1100



In summary, the dataset for bullying reveals a moderate range of scores with a central tendency around 2.617. The distribution is moderately spread, and there is a slight rightward skew with a flatter peak. The mode at 1 suggests a common bullying score in the dataset.

Stress Level

stress_level	
Mean	0.996364
Standard Error	0.024774
Median	1
Mode	0
Standard Deviation	0.821673
Sample Variance	0.675146
Kurtosis	-1.51891
Skewness	0.006728
Range	2
Minimum	0
Maximum	2
Sum	1096
Count	1100



In summary, the dataset for stress level reveals a moderate range of scores with a central tendency around 0.996. The distribution is moderately spread, symmetrical, and there is a mode at 0, suggesting a common stress level in the dataset.

• Inferential Statistics:

We used regression analysis to delve deeper into the relationship between academic performance and various influencing factors. We primarily relied on three key metrics to illuminate these connections: p-values, confidence intervals, and regression coefficients. These powerful tools allowed us to assess the statistical significance, precision, and magnitude of each factor's impact on academic outcomes.

For each factor, we are understanding the correlation between the factor and academic performance. Regression analysis was conducted to find the p-value and correlation coefficient.

Statistical Question:
H_o: r=0
H_1: r≠0
In this case, we determined that the hypothesized correlation coefficient for 16 factors should be zero.
D0 = 0
If we do not reject the hypothesis this indicates that factor has no impact on academic performance, and both the variables are independent. If we
reject the hypothesis this indicates that that factor has significant impact on academic performance, and both the variables are dependent. Factor can positively/negatively affect academic performance of student.

Anxiety level:

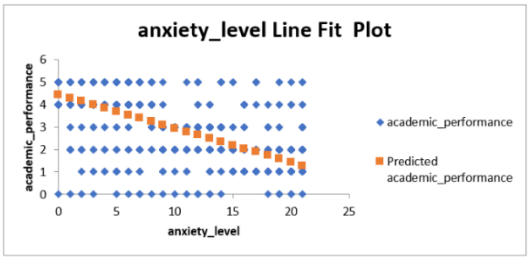
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.64960113
R Square	0.421981628
Adjusted R Square	0.421455199
Standard Error	1.075970141
Observations	1100

ANOVA

	df	SS	MS	F	Significance F
Regression	1	928.014323	928.014	801.594	7.487E-133
Residual	1098	1271.167495	1.15771		
Total	1099	2199.181818			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	4.434601978	0.067066283	66.1227	0	4.30300942	4.566194535	4.303009422	4.566194535
anxiety_level	-0.150210532	0.005305463	-28.312	7E-133	-0.16062052	-0.139800541	-0.160620523	-0.139800541



The regression analysis reveals a statistically significant relationship between the dependent variable and anxiety level, with a correlation coefficient of 0.65 and an R Square of 42.2%, indicating that the model explains a moderate portion of the variance in the academic performance The negative coefficient for anxiety level (-0.15) implies that as anxiety level increases, the dependent variable decreases by 0.15 units. The model is robust, as supported by a significant F-statistic (801.59), low p-values for both intercept and anxiety level coefficients, and narrow confidence intervals. The 95% confidence intervals for the intercept (4.30 to 4.57) and anxiety level (-0.16 to -0.14) indicate the range within which we can be 95% confident that the true population parameters lie, providing additional insight into the precision of the estimated coefficients. Overall, the findings suggest that anxiety level is a meaningful predictor of the dependent variable, and the confidence intervals enhance our understanding of the reliability of the model's predictions.

Self esteem:

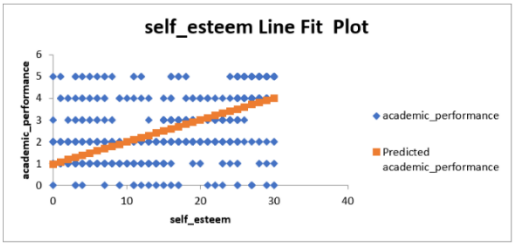
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.639045185
R Square	0.408378748
Adjusted R Square	0.407839931
Standard Error	1.088557266
Observations	1100

ANOVA

	df	SS	MS	F	Significance F
Regression	1	898.0991176	898.0991176	757.917103	2.6759E-127
Residual	1098	1301.082701	1.184956922		
Total	1099	2199.181818			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0.976060462	0.073049914	13.36155531	7.6867E-38	0.832727264	1.11939366	0.832727264	1.11939366
self_esteem	0.101065379	0.003671061	27.53029428	2.676E-127	0.093862293	0.10826847	0.093862293	0.108268466



The regression analysis reveals a statistically significant relationship between the dependent variable and self-esteem, with a correlation coefficient of 0.64 and an R Square of 40.8%, indicating that the model explains a substantial portion of the variance in the academic performance The positive coefficient for self-esteem (0.10) suggests that higher levels of self-esteem are associated with an increase in the academic performanceLow p-values for both intercept and self-esteem coefficients, and narrow confidence intervals. The 95% confidence intervals for the intercept (0.83 to 1.12) and self-esteem (0.09 to 0.11) indicate the range within which we can be 95% confident that the true population parameters lie. This provides a measure of the precision of the estimated coefficients, supporting the reliability of the model's predictions. Overall, the findings highlight the importance of self-esteem as a significant predictor of the dependent variable in this analysis.

Mental Health History :

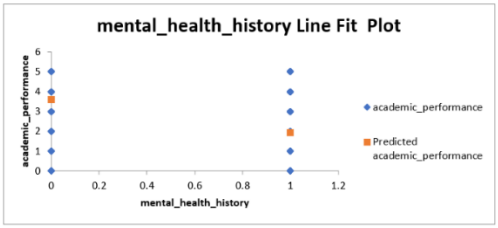
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.586193065
R Square	0.343622309
Adjusted R Square	0.343024516
Standard Error	1.146585127
Observations	1100

ANOVA

	df	SS	MS	F	Significance F
Regression	1	755.6879352	755.6879352	574.8173668	1.699E-102
Residual	1098	1443.493883	1.314657453		
Total	1099	2199.181818			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	3.589605735	0.048538819	73.95329792	0	3.494366414	3.684845055	3.494366414	3.684845055
mental_health_history	-1.657871417	0.069149	-23.97534915	1.699E-102	-1.793550528	-1.522192307	-1.793550528	-1.522192307



The regression analysis indicates a significant relationship between the dependent variable and mental_health_history, with a moderate positive linear association (Multiple R = 0.59). The model explains 34.4% of the variance in the dependent variable, and the negative coefficient for mental_health_history (-1.66) suggests that individuals with a mental health history, on average, score 1.66 units lower on the dependent variable compared to those without such a history. The model is statistically robust, supported by a significant F-statistic (574.82) and low p-value (1.699E-102). The 95% confidence interval for the intercept (3.49 to 3.68) and the mental_health_history coefficient (-1.79 to -1.52) indicates the range within which we can be 95% confident that the true population parameters lie. In practical terms, this suggests that individuals with a mental health history tend to have lower scores on dependent variable, and the confidence intervals provide a measure of the precision of these estimates

Depression :

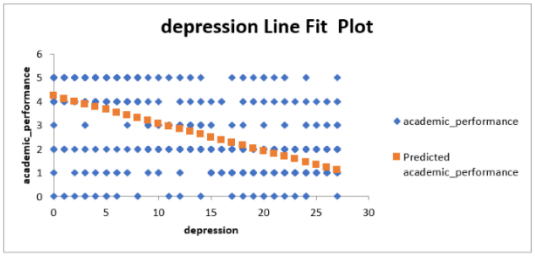
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.633174347
R Square	0.400909753
Adjusted R Square	0.400364134
Standard Error	1.095407027
Observations	1100

ANOVA

	df	SS	MS	F	Significance F
Regression	1	881.6734406	881.6734406	734.7789617	2.6465E-124
Residual	1098	1317.508378	1.199916555		
Total	1099	2199.181818			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	4.22810623	0.063035758	67.07472622	0	4.104422075	4.351790384	4.104422075	4.351790384
depression	-0.115916071	0.004276272	-27.10680656	2.6465E-124	-0.124306658	-0.107525483	-0.124306658	-0.107525483



The regression analysis reveals a statistically significant relationship between the dependent variable and depression, with a moderate positive linear association (Multiple R = 0.63). The model explains 40.1% of the variance in the dependent variable, and the negative coefficient for depression (-0.116) suggests that, on average, each one-unit increase in depression is associated with a decrease of 0.116 units in the academic performance The model is statistically robust, supported by a significant F-statistic (734.78) and an extremely low p-value (2.65E-124). The 95% confidence interval for the intercept (4.10 to 4.35) and the depression coefficient (-0.124 to -0.108) indicates the range within which we can be 95% confident that the true population parameters lie. In practical terms, this implies that higher levels of depression are associated with lower scores on the dependent variable, and the confidence intervals provide a measure of the precision of these estimates, indicating the range for the true effects of depression on the dependent variable

Headache:

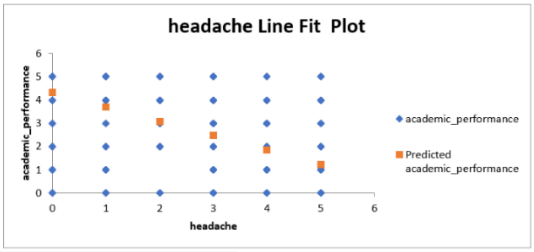
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.622058546
R Square	0.386956835
Adjusted R Square	0.386398508
Standard Error	1.108089719
Observations	1100

ANOVA

	df	SS	MS	F	Significance F
Regression	1	850.9884364	850.9884364	693.064746	8.3114E-119
Residual	1098	1348.193382	1.227862825		
Total	1099	2199.181818			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	4.338761687	0.06822622	63.59375734	0	4.204893188	4.472630187	4.204893188	4.472630187
headache	-0.624370372	0.023716761	-26.32612287	8.311E-119	-0.67090567	-0.57783508	-0.67090567	-0.57783508



The regression analysis reveals a statistically significant relationship between the dependent variable and headache, with a moderate positive linear association (Multiple R = 0.62). The model explains 38.7% of the variance in the dependent variable, and the negative coefficient for headache (-0.624) suggests that, on average, each one-unit increase in headache is associated with a decrease of 0.624 units in the academic performance The model is statistically robust, supported by a significant F-statistic (693.06) and an extremely low p-value (8.31E-119). The 95% confidence interval for the intercept (4.20 to 4.47) and the headache coefficient (-0.671 to -0.578) indicates the range within which we can be 95% confident that the true population parameters lie. In practical terms, this implies that higher levels of headache are associated with lower scores on the dependent variable, and the confidence intervals provide a measure of the precision of these estimates, showing the range for the true effects of headache on the academic performance

Blood Pressure:

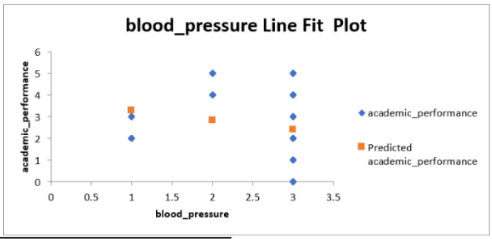
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.262785392
R Square	0.069056162
Adjusted R Square	0.068208308
Standard Error	1.365498222
Observations	1100

ANOVA

	df	SS	MS	F	Significance F
Regression	1	151.8670563	151.8670563	81.44816366	7.86613E-19
Residual	1098	2047.314762	1.864585393		
Total	1099	2199.181818			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	3.745714286	0.115405663	32.45693653	1.4359E-162	3.519273733	3.972154838	3.519273733	3.972154838
blood_pressure	-0.445952381	0.049413753	-9.024863637	7.86613E-19	-0.542908433	-0.348996329	-0.542908433	-0.348996329



The regression analysis suggests a statistically significant relationship between the dependent variable and blood_pressure,with a modest Multiple R (0.26) and an R Square of 6.9%. The negative coefficient for blood_pressure (-0.446) indicates that, on average, each one-unit increase in blood pressure is associated with a decrease of 0.446 units in the academic performance Despite the limited explained variance, the model is statistically robust, supported by a significant F-statistic (81.45) and a very low p-value (7.87E-19). The 95% confidence interval for the intercept (3.52 to 3.97) and the blood_pressure coefficient (-0.543 to -0.349) indicates the range within which we can be 95% confident that population parameters lie. In practical terms, this implies that while there is a statistically significant negative association between blood pressure and the dependent variable, caution is needed in making predictions due to the limited overall explanatory power of the model. The confidence intervals provide a measure of the precision of these estimates, showing the range for the true effects of blood pressure on the academic performance.

Sleep Quality:

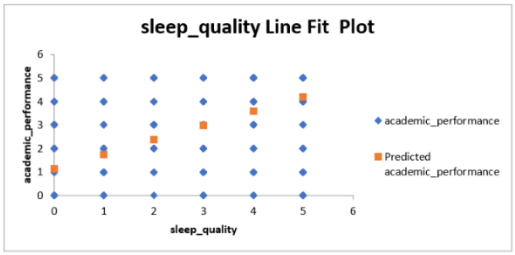
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.671326304
R Square	0.450679007
Adjusted R Square	0.450178715
Standard Error	1.048920314
Observations	1100

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	991.1250778	991.1250778	900.83131	5.2325E-145
Residual	1098	1208.05674	1.100233825		
Total	1099	2199.181818			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	1.141296142	0.062887048	18.14834986	1.4126E-64	1.017903776	1.26468851	1.01790378	1.26468851
sleep_quality	0.613319974	0.020434564	30.01385196	5.232E-145	0.573224767	0.65341518	0.57322477	0.65341518



The regression analysis reveals a highly significant and positive relationship between the dependent variable and sleep quality, with a strong Multiple R value (0.67) and a substantial R Square of 45.1%, indicating that half of the variance in the dependent variable is explained by sleep quality. The positive coefficient for sleep_quality (0.613) suggests that, on average, each one-unit increase in sleep quality corresponds to a 0.613-unit increase in the academic performance The model is statistically robust, supported by a high F-statistic (900.83) and an extremely low p-value (5.23E-145). The 95% confidence interval for the intercept (1.02 to 1.26) and the sleep_quality coefficient (0.57 to 0.65) indicates the range within which we can be 95% confident that the true population parameters lie. This implies that higher sleep quality is strongly associated with higher scores on the dependent variable, and the confidence intervals provide a measure of the precision of these estimates, indicating the range for the true effects of sleep quality on the academic performance

Breathing Problems:

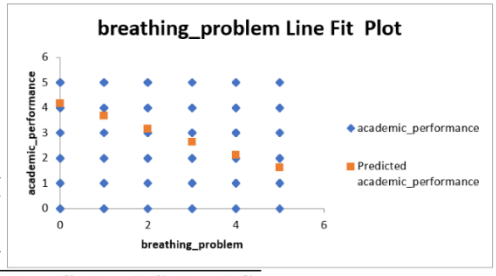
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.507250875
R Square	0.25730345
Adjusted R Square	0.256627042
Standard Error	1.219649666
Observations	1100

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	565.85707	565.85707	380.396528	5.6318E-73
Residual	1098	1633.324748	1.487545308		
Total	1099	2199.181818			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	4.183353226	0.081137807	51.55861831	1.601E-295	4.024150555	4.342555897	4.024150555	4.342555897
breathing_problem	-0.5122775	0.026265581	-19.50375676	5.6318E-73	-0.5638139	-0.4607411	-0.5638139	-0.4607411



The regression analysis demonstrates a statistically significant relationship between the dependent variable and the independent variable, breathing_problem, with a moderate Multiple R value (0.51) and a substantial R Square of 25.7%, indicating that about a quarter of the variance in the dependent variable is explained by the severity of breathing problems. The negative coefficient for breathing_problem (-0.512) suggests that, on average, each one-unit increase in the severity of breathing problems corresponds to a decrease of 0.512 units in the academic performance The model is statistically robust, supported by a significant F-statistic (380.40) and an extremely low p-value (5.63E-73). The 95% confidence interval for the intercept (4.02 to 4.34) and the breathing_problem coefficient (-0.564 to -0.461) indicates the range within which we can be 95% confident that the true population parameters lie. Practically, this implies that the presence of breathing problems is negatively associated with scores on the dependent variable, and the confidence intervals provide a measure of the

precision of these estimates, showing the range for the true effects of breathing problems on academic performance.

Noise Level :

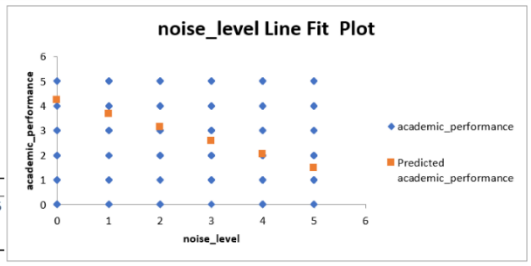
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.513729808
R Square	0.263918316
Adjusted R Square	0.263247932
Standard Error	1.214206082
Observations	1100

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	580.4043614	580.4043614	393.6822731	4.09347E-75
Residual	1098	1618.777457	1.474296409		
Total	1099	2199.181818			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	4.222245878	0.081714901	51.67045196	2.9675E-296	4.061910875	4.382580881	4.061910875	4.382580881
noise_level	-0.547175863	0.027577443	-19.8414282	4.09347E-75	-0.601286306	-0.493065421	-0.601286306	-0.493065421



The regression analysis indicates a significant relationship between the predictor variable "noise_level" and the response variable. The model has an R-squared value of 0.26, indicating that 26% of the variance in the response variable is explained by "noise_level." Both the intercept and "noise_level" coefficients are statistically significant ($p < 0.05$), with a negative coefficient of -0.547. This negative coefficient implies an inverse relationship, suggesting that higher noise levels are associated with lower values in the response variable. The 95% confidence intervals for the coefficients provide a range within which we can be 95% confident about their true values, indicating the precision of our coefficient estimates. Overall, the findings suggest that elevated noise levels are a meaningful predictor of the response variable, with higher levels associated with a significant decrease in the outcome.

Living Conditions:

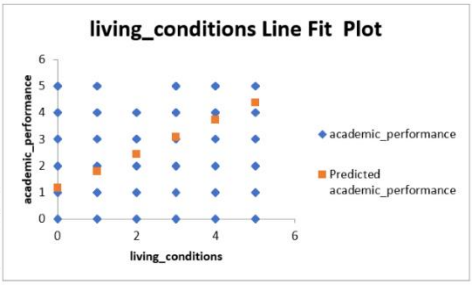
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.507220614
R Square	0.257272752
Adjusted R Square	0.256596315
Standard Error	1.219674872
Observations	1100

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	565.7895577	565.7895577	380.3354218	5.7614E-73
Residual	1098	1633.39226	1.487606795		
Total	1099	2199.181818			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	1.158350393	0.090580217	12.78811673	5.24038E-35	0.980620515	1.336080271	0.980620515	1.336080271
living_conditions	0.641088292	0.03287263	19.50219018	5.7614E-73	0.57658802	0.705588563	0.57658802	0.705588563



The regression analysis indicates a significant relationship between the predictor variable "living_conditions" and the response variable. The model has an R-squared value of 0.26, suggesting that 26% of the variance in the response variable is explained by "living_conditions." Both the intercept and "living_conditions" coefficients are statistically significant ($p < 0.05$), with a positive coefficient of 0.641. This positive coefficient implies a direct relationship, indicating that better living conditions are associated with higher values in the response variable. The 95% confidence intervals for the coefficients provide a range within which we can be 95% confident about their true values, indicating the precision of our coefficient estimates. Overall, the findings suggest that improved living conditions are a

meaningful predictor of the response variable, with higher levels associated with a significant increase in the outcome.

Safety:

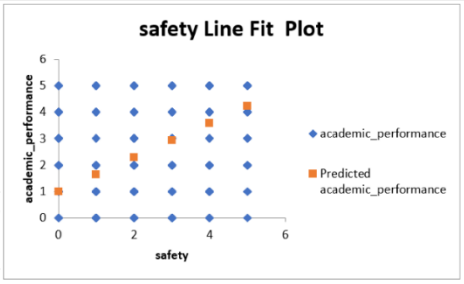
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.642846165
R Square	0.413251191
Adjusted R Square	0.412716812
Standard Error	1.084065457
Observations	1100

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	908.8145066	908.8145066	773.3288957	2.8388E-129
Residual	1098	1290.367312	1.175197916		
Total	1099	2199.181818			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	1.002542275	0.07155694	14.01041287	3.74934E-41	0.86213848	1.142946069	0.86213848	1.142946069
safety	0.646696612	0.023255114	27.8087917	2.8388E-129	0.601067128	0.692326095	0.601067128	0.692326095



The regression analysis indicates a significant relationship between the predictor variable "safety" and the response variable. The model has an R-squared value of 0.41, indicating that 41% of the variance in the response variable is explained by "safety." Both the intercept and "safety" coefficients are statistically significant ($p < 0.05$), with a positive coefficient of 0.647. This positive coefficient implies a direct relationship, suggesting that a sense of safety is associated with higher values in the response variable. The 95% confidence intervals for the coefficients provide a range within which we can be 95% confident about their true values, indicating the precision of our coefficient estimates. Overall, the findings suggest that a perceived sense of safety is a meaningful predictor of the response variable, with higher levels of safety associated with a significant increase in the outcome.

Basic Needs:

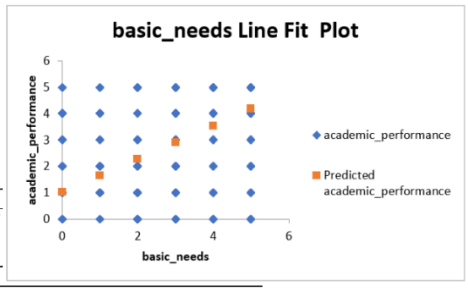
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.639387458
R Square	0.408816321
Adjusted R Square	0.408277903
Standard Error	1.088154634
Observations	1100

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	899.0614206	899.0614206	759.290787	1.7817E-127
Residual	1098	1300.120398	1.184080508		
Total	1099	2199.181818			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	1.02358054	0.071455397	14.32474788	8.52648E-43	0.883375987	1.163785093	0.883375987	1.163785093
basic_needs	0.630839805	0.022893649	27.55523157	1.7817E-127	0.585919561	0.675760049	0.585919561	0.675760049



The regression analysis indicates a significant relationship between the predictor variable "basic_needs" and the response variable. The model has an R-squared value of 0.41, suggesting that 41% of the variance in the response variable is explained by "basic_needs." Both the intercept and "basic_needs" coefficients are statistically significant ($p < 0.05$), with a positive coefficient of 0.631. This positive coefficient implies a direct relationship, suggesting that meeting basic needs is associated with higher values in the response variable. The 95% confidence intervals for the coefficients provide a range within which we can be 95% confident about their true values, indicating the precision of our coefficient estimates. Overall, the

findings suggest that fulfilling basic needs is a meaningful predictor of the response variable, with higher levels of meeting basic needs linked to a significant increase in the outcome.

Social Support:

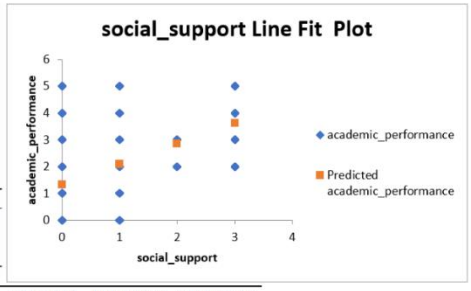
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.567501207
R Square	0.32205762
Adjusted R Square	0.321440186
Standard Error	1.165267919
Observations	1100

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	708.2632617	708.2632617	521.6066686	8.94508E-95
Residual	1098	1490.918556	1.357849323		
Total	1099	2199.181818			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	1.330987719	0.072245611	18.42309444	2.96847E-66	1.189232665	1.472742774	1.189232665	1.472742774
social_support	0.766141792	0.033545756	22.83870987	8.94508E-95	0.700320762	0.831962821	0.700320762	0.831962821



The regression analysis indicates a significant relationship between the predictor variable "social_support" and the response variable. The model has an R-squared value of 0.32, suggesting that 32% of the variance in the response variable is explained by "social_support." Both the intercept and "social_support" coefficients are statistically significant ($p < 0.05$), with a positive coefficient of 0.766. This positive coefficient implies a direct relationship, suggesting that higher levels of social support are associated with higher values in the response variable. The 95% confidence intervals for the coefficients provide a range withinbre which we can be 95% confident about their true values, indicating the precision of our coefficient estimates. Overall, the findings suggest that social support is a meaningful predictor of the response variable, with higher levels of social support linked to a significant increase in the outcome.

Peer Pressure:

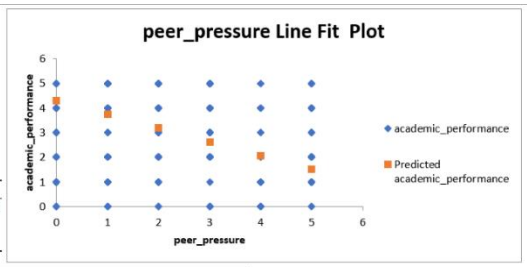
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.562947656
R Square	0.316910063
Adjusted R Square	0.316287941
Standard Error	1.169683441
Observations	1100

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	696.9428495	696.9428495	509.401809	5.73568E-93
Residual	1098	1502.238969	1.368159352		
Total	1099	2199.181818			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	4.300607231	0.076331153	56.34144209	0	4.150835826	4.450378637	4.150835826	4.450378637
peer_pressure	-0.558732698	0.024755622	-22.56993153	5.7357E-93	-0.60730637	-0.51015903	-0.60730637	-0.51015903



The regression analysis indicates a significant relationship between the predictor variable "peer_pressure" and the response variable. The model has an R-squared value of 0.32, suggesting that 32% of the variance in the response variable is explained by "peer_pressure." Both the intercept and "peer_pressure" coefficients are statistically significant ($p < 0.05$), with a negative coefficient of -0.559. This negative coefficient implies an inverse relationship, suggesting that higher levels of peer pressure are associated with lower values in the response variable. The 95% confidence intervals for the coefficients provide a range within which we can be 95% confident about their true values, indicating the precision of our coefficient estimates. Overall, the findings suggest that peer pressure is a meaningful predictor of the

response variable, with higher levels of peer pressure linked to a significant decrease in the outcome.

Extracurricular Activities :

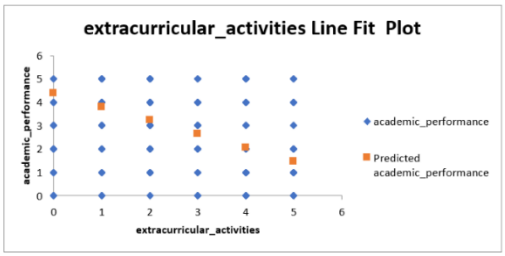
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.588612215
R Square	0.346464339
Adjusted R Square	0.345869134
Standard Error	1.144100152
Observations	1100

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	761.9380752	761.9380752	582.0919456	1.5625E-103
Residual	1098	1437.243743	1.308965158		
Total	1099	2199.181818			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	4.398166688	0.075689244	58.10821232	0	4.249654787	4.546678588	4.249654787	4.546678588
extracurricular_activities	-0.587379552	0.024345743	-24.12658172	1.5625E-103	-0.635148988	-0.539610116	-0.635148988	-0.539610116



The regression analysis indicates a significant relationship between the predictor variable "extracurricular_activities" and the response variable. The model has an R-squared value of 0.35, signifying that 35% of the variance in the response variable is explained by "extracurricular_activities." Both the intercept and "extracurricular_activities" coefficients are statistically significant ($p < 0.05$), with a negative coefficient of -0.587. This negative coefficient implies an inverse relationship, suggesting that higher levels of involvement in extracurricular activities are associated with lower values in the response variable. The 95% confidence intervals for the coefficients provide a range within which we can be 95% confident about their true values, indicating the precision of our coefficient estimates. Overall, the findings suggest that participation in extracurricular activities is a meaningful predictor of the response variable, with higher levels of involvement linked to a significant decrease in the outcome.

Bullying:

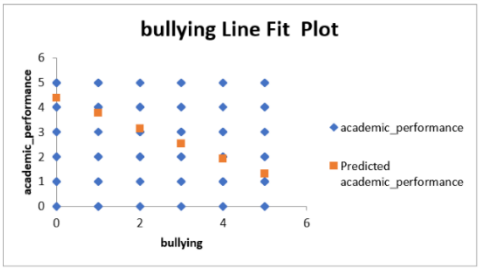
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.666228726
R Square	0.443860716
Adjusted R Square	0.443354214
Standard Error	1.055409952
Observations	1100

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	976.1304156	976.1304156	876.325553	4.6047E-142
Residual	1098	1223.051403	1.113890166		
Total	1099	2199.181818			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	4.383895091	0.063046325	69.53450685	0	4.260190203	4.507599978	4.260190203	4.507599978
bullying	-0.615590344	0.020795007	-29.60279637	4.605E-142	-0.65639278	-0.5747879	-0.65639278	-0.5747879



The regression analysis reveals a significant relationship between the predictor variable "bullying" and the response variable. The model has an R-squared value of 0.44, indicating that 44% of the variance in the response variable is explained by "bullying." Both the intercept and "bullying" coefficients are statistically significant ($p < 0.05$), with a negative coefficient for "bullying" (-0.616). This negative coefficient implies an inverse relationship, suggesting that higher levels of bullying are associated with lower values in the response variable. The 95% confidence intervals for the coefficients provide a range within which we can be 95% confident about their true values. In this context, the confidence interval indicates the precision

of our coefficient estimates. Overall, the findings suggest that "bullying" is a meaningful predictor of the response variable, with higher levels of bullying associated with a significant decrease in the outcome.

Stress Level:

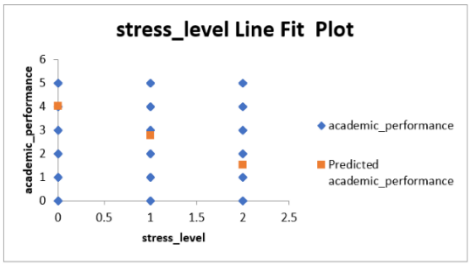
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.720922361
R Square	0.51972905
Adjusted R Square	0.519291645
Standard Error	0.980782159
Observations	1100

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1142.978678	1142.978678	1188.209484	4.5633E-177
Residual	1098	1056.20314	0.961933643		
Total	1099	2199.181818			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	4.009355734	0.046491995	86.23754939	0	3.918132541	4.100578926	3.918132541	4.100578926
stress_level	-1.241141703	0.036005999	-34.47041461	4.5633E-177	-1.311790041	-1.170493366	-1.311790041	-1.170493366



The regression analysis demonstrates a significant relationship between stress_level and the response variable, with an R-squared value of 0.52, indicating that 52% of the variance is explained by stress_level. Both the intercept and stress_level coefficients are statistically significant ($p < 0.05$), revealing a negative association, where higher stress levels correspond to lower values in the response variable. The 95% confidence intervals for the coefficients provide a range within which we can be 95% confident about their true values. In this context, the confidence interval indicates the precision of our coefficient estimates. The results suggest that stress_level is a meaningful predictor of the response variable, with higher stress levels associated with a significant decrease in the outcome.

• Conclusion:

Physiological factors

Physiological factors paint a complex picture of their influence on the academic performance:

- **Breathing problems:** Exert a significant negative impact, with each unit increase linked to a 0.512 unit decrease in academic performance. This association is statistically robust and holds true with 95% confidence.
- **Sleep quality**: Strongly and positively influences academic performance, explaining nearly half of its variance. Each unit improvement in sleep quality corresponds to a 0.613 unit increase, a highly significant and precise finding.
- **Blood pressure**: Though statistically relevant, holds limited explanatory power (6.9%). Each unit increase in blood pressure associates with a 0.446 unit decrease in the academic performance, but caution is warranted due to the model's weakness.
- **Headache**: Demonstrates a statistically significant and moderately positive linear association, explaining 38.7% of the variance. Each unit increase in headache leads to a 0.624 unit decrease in the academic performance. This relationship is robust and holds true with 95% confidence.

Overall, while breathing problems and headaches negatively impact the academic performance sleep quality significantly boosts it. Blood pressure's influence, while statistically relevant, requires cautious interpretation due to the model's limitations.

Psychological factors

Psychological factors paint a nuanced picture of their influence on the academic performance:

- **Anxiety level**: Exhibits a significant negative association, explaining 42.2% of the variance. Higher anxiety levels correspond to lower scores on the academic performance, with a robust and precise relationship.
- **Self-esteem**: Demonstrates a significant positive association, explaining 40.8% of the variance. Increased self-esteem is linked to higher scores on the academic performance, supported by a robust and reliable model.
- **Mental health history**: Reveals a significant negative association, explaining 34.4% of the variance. Individuals with a mental health history score lower on the academic performance compared to those without, with a statistically robust finding.
- **Depression:** Shows a significant negative association, explaining 40.1% of the variance. Higher depression levels correspond to lower scores on academic performance, supported by a robust and precise model.

Overall, anxiety, depression, and mental health history negatively impact academic performance, while self-esteem exerts a positive influence. The models for all factors are statistically robust, providing reliable and precise estimates of their effects. These findings highlight the complex interplay between psychological well-being and academic performance, suggesting potential avenues for intervention and support.

Environmental factors

Environmental factors drastically can alter the academic performance:

Environmental factors play a significant role in academic performance. These findings suggest that interventions aimed at improving the physical and social environment of students can have a positive impact on their academic outcomes.

- **Noise Levels:** Noise levels have a negative coefficient of -0.547, meaning that for every one-unit increase in noise level, there is a 0.547 decrease in academic performance. This is likely because noise can make it difficult to concentrate and learn. The R-squared values for noise level, safety, and basic needs are 0.26. This means that this factor explains 26% of the variance in academic performance, respectively.
- **Safety, living conditions, and basic needs**: They all have positive coefficients, meaning that for every one-unit increase in these factors, there is an increase in academic performance. This is likely because these factors can reduce stress and anxiety, which can interfere with learning. The R-squared values for living conditions, safety, and basic needs are 0.26, 0.41, and 0.41 respectively. This means that these factors explain 26%, 41%, and 26% of the variance in academic performance, respectively.

These are all relatively high R-squared values, suggesting that these factors are important predictors of academic performance. Overall, the findings of this study suggest that environmental factors are

important for academic performance and that interventions aimed at improving these factors can have a positive impact on student outcomes.

Social Factors
Social Factors can influence a student's ability to take key academic decisions:

Social support has a positive impact on academic performance, while peer pressure, extracurricular activities, and bullying have a negative impact. These findings suggest that interventions aimed at improving social support and reducing peer pressure, extracurricular activities, and bullying can have a positive impact on student outcomes.

- **Social Support**: Increased social support can greatly boost the morale of students. It has a positive coefficient of 0.766. The R-squared value of 0.32 suggests a 32% variation in academic performance. Overall, the findings suggest that social support is a meaningful predictor of the response variable, with higher levels of social support linked to a significant increase in the outcome.
- **Peer Pressure:** Increased social support can greatly boost the morale of students. It has a negative coefficient of 0.559. The R-squared value of 0.32 suggests a 32% variation in academic performance. Overall, the findings suggest that social support is a meaningful predictor of the response variable, with higher levels of social support linked to a significant increase in the outcome. Overall, the findings suggest that peer pressure is a meaningful predictor of the response variable, with higher levels of peer pressure linked to a significant decrease in the outcome.
- **Extracurricular Activities**: Extracurricular Activities have a negative coefficient of 0.587. The R-squared value of 0.35 suggests a 35% variation in academic performance. Overall, the findings suggest that participation in extracurricular activities is a meaningful predictor of the response variable, with higher levels of involvement linked to a significant decrease in the outcome.
- **Bullying**: Bullying is a significant predictor of academic performance. Bullying has a negative coefficient of 0.616. The model explains 44% of the variance in academic performance because of R-squared value of 0.44. Higher levels of bullying are associated with lower values of the response variable.

Overall, the findings of this study suggest that social factors are important for academic performance and that interventions aimed at improving these factors can have a positive impact on student outcomes.

• Future Improvements:

- Longitudinal studies: Conduct longitudinal studies to track changes in the dependent variable over time and evaluate the effectiveness of interventions.
- Mediation and moderation analysis: Investigate potential mediating factors (e.g., coping mechanisms) and moderating factors (e.g., personality traits) that might influence the relationships observed.
- Multi-level analysis: Consider incorporating data at individual, social, and environmental levels to gain a more comprehensive understanding of the factors at play.
- Personalized interventions: Develop personalized intervention plans that address the unique needs and challenges of everyone.

- **References:**

https://www.kaggle.com/datasets/rxnach/student-stress-factors-a-comprehensive-analysis/data