Statistics 810: Analysis on Student AI Usage

Archisha Bhatt, Julia Bunescu, Sai Keertana Gubbala

December 11, 2025

Abstract

This study analyzes 9,551 student-AI interaction sessions to address three primary questions: (1) What factors predict student satisfaction with AI educational tools? (2) How does usage differ between disciplines, levels, and associated tasks? (3) What characteristics influence repeat usage of an AI tool among students?

1 Introduction

With the boom of artificial intelligence (AI) in the past years, educational tools powered by AI, as well as the general use of it have become increasingly common within universities and schools. This still comes with its own criticism of whether it aids learning or could negatively impact student's cognitive development (Edwards [2018]). The applications of AI have revolutionized the learning landscape for students across the world, whether that is by offering personalized lessons to entire content creation (Tan [2025]). When discussing AI, we are referring exclusively to proprietary Large Language Models (LLMs).

Now, as students across different grade levels and disciplines increasingly turn to AI for assistance with concepts or coursework, it is important to understand how and where they are using these tools. To explore further, we focus on analyzing the patterns in student engagement with AI, looking at trends in usage frequency, satisfaction, and differences between subjects and task type.

2 Data

The data for this project was sourced from the Data Usage AI in School dataset (Baariq [2025]), which contains 9,551 sessions of student using an AI tool, with 13 features, collected through passive logging of user activity on an AI educational assistant platform. This dataset contains anonymized session-level data that reveals how students are incorporating AI into their learning, offering insights into the future of what could be modern education.

Key metrics include assignment type, student level, and duration of each session. The data also has no missing values and is not synthetic. As the data came in a ready-to-use format with minimal data cleaning, we could proceed directly with analysis.

2.1 Dataset Structure

Table-1 shows an example of the data available in the dataset (non-exhaustive):

Table 1: Sample of Student AI Usage Dataset

Discipline	FinalOutcome	Satisfaction	Year	Quarter	Month	Day	Session (min)
Biology	Assignment Completed	3.00	2024	Qtr 2	June	24	0.84
Biology	Assignment Completed	2.00	2024	Qtr 2	June	24	16.03
Biology	Assignment Completed	3.70	2024	Qtr 2	June	24	13.49
Biology	Assignment Completed	4.90	2024	Qtr 2	June	25	9.56
Biology	Assignment Completed	3.80	2024	Qtr 2	June	26	7.78

2.2 Descriptive Statistics

Table-2 shows descriptive statistics from a subset of our data, using our continuous, numerical features. Session duration is the only variable that has the most variation compared to the others.

Table 2: Descriptive statistics for key numeric variables (n = 9,551)

	count	mean	std	min	25%	50%	75%	max
Satisfaction	9551.00	3.58	1.40	1.00	2.60	3.60	4.50	14.20
Session (min)	9551.00	20.78	14.91	0.03	9.89	17.24	27.97	142.63
Prompts	9551.00	5.87	4.91	1.00	2.00	4.00	8.00	50.00
Assistance	9551.00	3.64	1.29	1.00	3.00	4.00	4.00	14.00

2.3 Categorical Variables Exploration

Figure-1 explores and evaluates how satisfaction, session length, prompts, and requested assistance levels to spot skewness, outliers, or dominant behaviors. We see that many of the distributions are right-skewed and slightly bimodal.

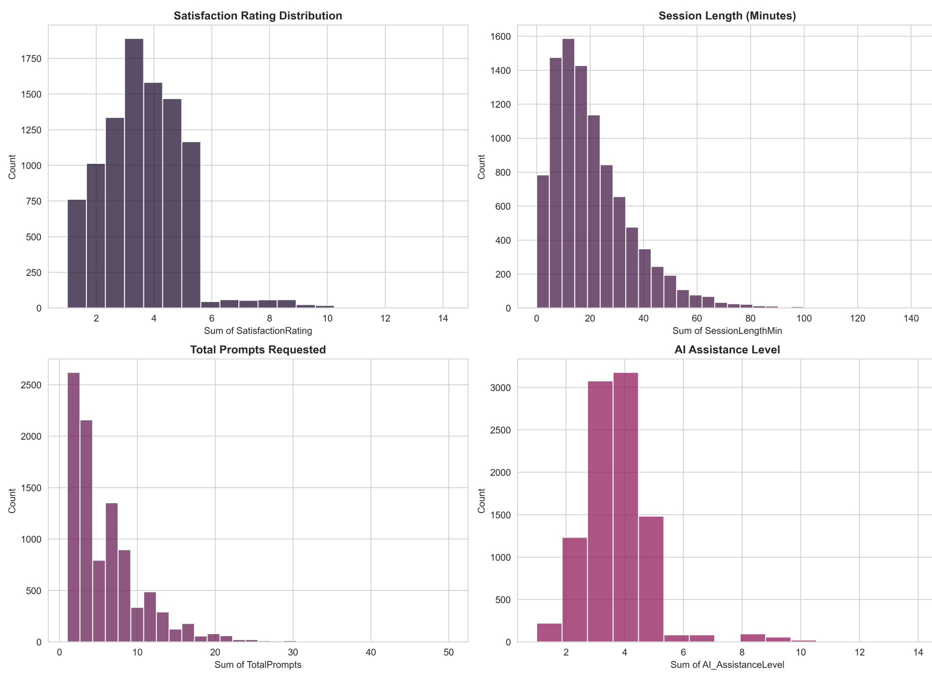

Figure 1: Distribution of Usage Metrics

We also investigated how AI tool usage is distributed across key student characteristics (Figure-2), including subject areas, educational level, task type, and academic outcomes, to better understand which traits contribute to heavy use of AI support. Our findings show distinct patterns in adoption and usage. With the distribution of disciplines, we saw nearly uniform adoption rates of AI tools across different fields of study, indicating that approximately equal proportions of students in each discipline engage in these technologies. The balanced distribution demonstrates that AI tools serve educational needs across a variety of academic fields.

However, when examining usage by educational level, a clear pattern emerges: Undergraduate students demonstrate significantly higher rates of AI tool consumption compared to high school and graduate students, representing a disproportionate share of overall usage. This can raise important questions about the specific circumstances or needs that drive undergraduates toward seeking AI assistance. Additionally, our analysis of task categories reveals that writing makes up the primary group in which students seek AI support, followed by studying, and homework help. Lastly, the outcomes due to utilizing the AI tool was positive, resulting in an assignment being completed or drafted, rather than leaving a student confused or giving up.

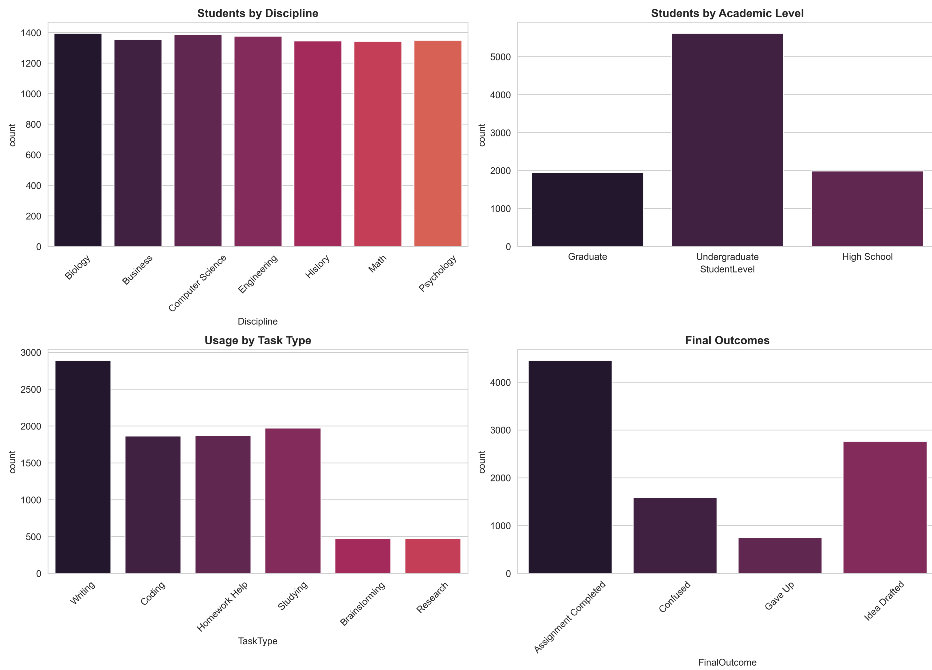

Figure 2: Categorical Insights

2.4 Sanity Check

To check the stability of our satisfaction variable's estimates, we performed bootstrap resampling with a 1,000 iterations. Figure-3 shows the resulting distribution of mean satisfaction scores. The bootstrap distribution hugs the observed mean of 3.578 tightly, and the 95% confidence interval is hardly wider than ± 0.03. That stability indicates that sampling noise is not driving the cohort average and future splits should expect similar central values.

3 Methods

To better understand how students use AI tools, we conducted multiple types of analysis. We began by examining basic summary statistics and distributions, and then looked at relationships between different metrics and temporal dynamics. Our analysis included examining one variable at a time (univariate), comparing two variables together (bivariate), and analyzing multiple variables simultaneously (multivariate) to understand student AI engagement patterns.

3.1 Statistical Testing

We used several statistical tests to determine whether patterns we observed were meaningful or simply due to chance. We began with creating visualizations such as violin plots, count plots, and scatter plots to investigate potential patterns and relationships among variables. These graphics helped guide our selection on what tests would be most appropriate.

To compare satisfaction differences between students who were repeat users of the AI tool versus those who were not, we applied Welch's t-tests (Lu [2010]). This test allowed us to determine whether the two groups differed significantly in their satisfaction levels.

Additionally, we also wanted to compare whether the mean satisfaction differs across student levels or task types, so we used one-way Analysis of Variance (ANOVA) tests (Fisher [1925]). For further

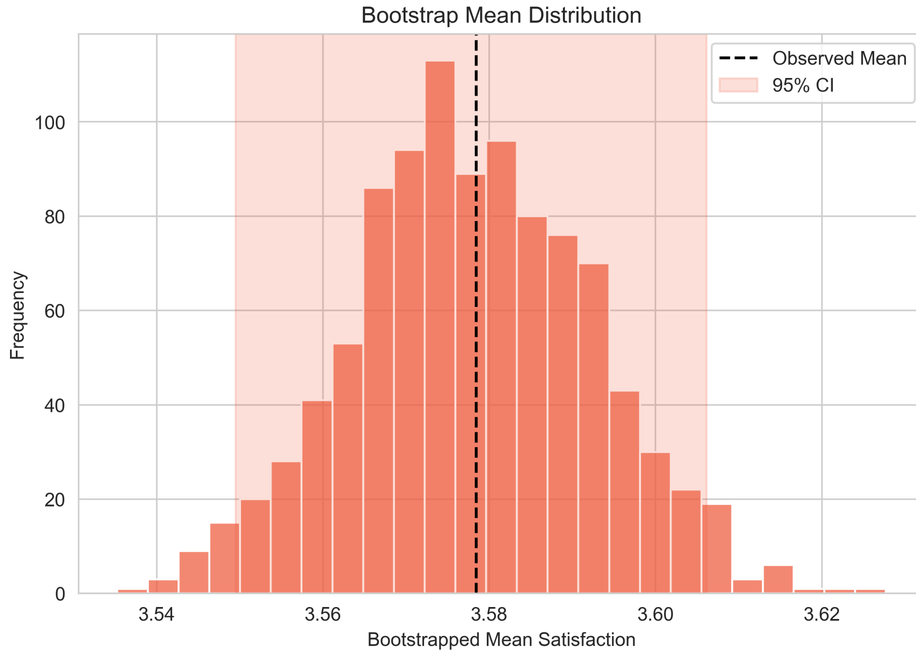

Figure 3: Bootstrapping of Satisfaction Mean

examining, we conducted post-hoc tests such as the Tukey HSD to identify exactly which pairs of groups differ significantly (Keselman [1977]).

Our scatter plots revealed several relationships that appear to trend linearly. To quantify this statistically, we calculated Pearson's R (Pearson [1909]) between AI assistance level and satisfaction, as well as between number of prompts to session length pairings. Based on the correlations results, we conducted a simple linear regression to model the relationships better.

Collectively, these methods allowed us to move beyond simple visualizations of the data and make statistical inferences to answer our research questions about AI usage patterns.

3.2 Reproducibility

All analyses were conducted in Python (version 3.12.3) using these libraries: Pandas (version 2.2.2) for data manipulation, NumPy (version 1.26.4), SciPy (version 1.16.3) and StatsModel (version 0.14.2) for statistical testing, Seaborn (version 0.13.2) and Matplotlib (version 3.9.2) for creation of visualizations. The complete analysis pipeline is documented in a Jupyter notebook to ensure reproducibility.

4 Analysis & Results

We looked at the dataset from a few angles and kept it grounded in how students actually use AI. Using a mix of common sense and our own consumer experience, we spotted the patterns behind their satisfaction and identified ideas that could shape a future modeling round. In the following sections, we walk through each angle to show what we found and why it matters.

4.1 Satisfaction Drivers

The first perspective we explored was the satisfaction viewpoint. The feature distribution (Figure-1) indicates that most scores cluster around 3 and 4, reflecting a neutral to slightly negative experience. There are also a few exceptional cases in which satisfaction exceeds the nominal upper bound of 10. Since we did not control the data collection procedures, nor did we have enough information on how the data was collected, we chose to retain these values in the analysis.

To further investigate satisfaction, we examined how scores vary across student levels and task types (Figure-4). We observed that values above 10 occur primarily among undergraduate students and that tasks such as writing, coding, and homework display longer tails in their distributions. To assess whether mean satisfaction differs by student level or task type, we complemented the visual analysis with simple one-way ANOVA tests (Fisher [1925]). These tests revealed statistically significant differences between groups, so we applied Tukey post-hoc comparisons (Keselman [1977]) to determine which pairs of groups differ significantly.

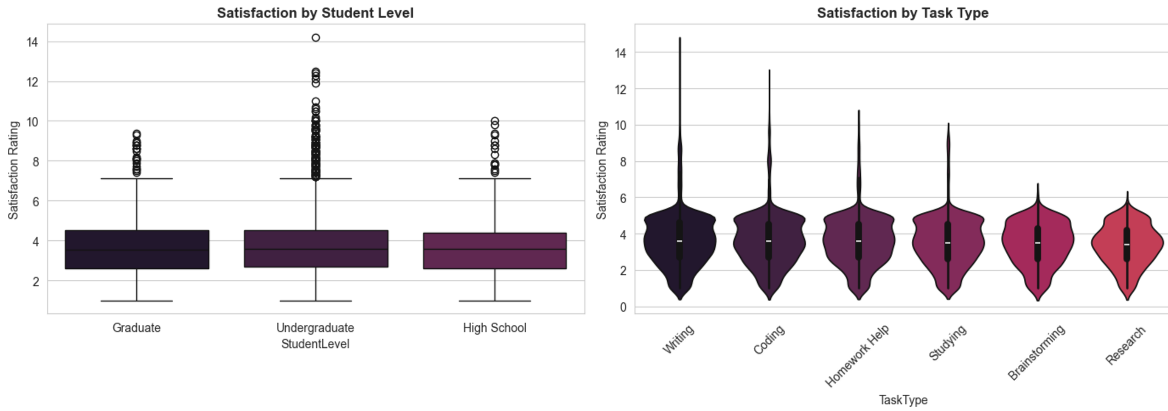

Figure 4: Satisfaction patterns across student segments and task types

Our significant results (Table-3) show, first, that graduate and high-school students report slightly higher satisfaction than undergraduates, and that these differences are large enough to be considered meaningful. Second, writing-intensive sessions receive the lowest satisfaction scores, suggesting that students are more satisfied with brainstorming, research, and studying sessions; research tasks also yield higher satisfaction than coding and homework-help sessions.

Table 3: Pairwise differences and adjusted p-values

Pair	Difference (diff)	Adjusted p-value (p-adj)
Brainstorming vs Writing	0.24	0.005
Coding vs Research	-0.26	0.003
Homework Help vs Research	-0.24	0.009
Research vs Writing	0.31	0.000
Studying vs Writing	0.14	0.012

We confirmed the modest gap between graduate and undergraduate satisfaction (Figure-5) by bootstrapping the difference in means: the 95% confidence interval stayed entirely below zero, so a real (though small) separation exists. Undergraduates still hold roughly a tenth-of-a-point advantage, and while the error bars mostly overlap, the lead persists even if the effect size remains modest.

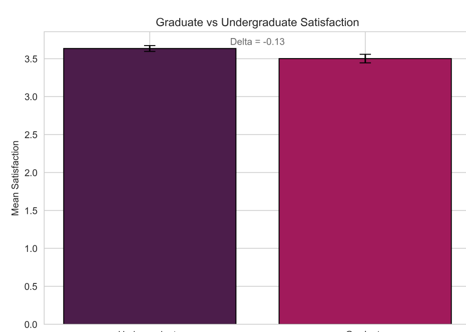

Figure 5: Graduate vs Undergraduate Satisfaction

4.2 Workload Patterns by Task Type

We then examined how engagement depth—measured by session length and prompt volume—varies across task types. Looking at the box plots (Figure-6) and confirming with a one-way ANOVA (Fisher [1925]), we found no statistically meaningful differences in session length between tasks, so the workload profile stays consistent regardless of task category.

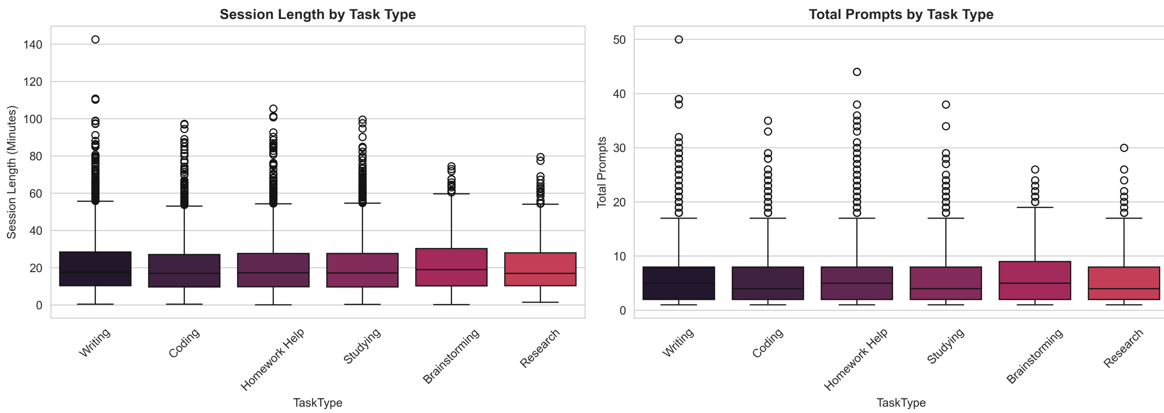

Figure 6: Workload Distribution by Task Type

4.3 Cross-Metric Relationships

Moving into multivariate analysis, we explored how the numerical features—counts and approximately continuous measures—relate to one another while using our categorical fields to hunt for segmentation patterns (see Figure-7). None of the categories delivered a strong split, but we did uncover potentially meaningful connections among the numeric variables themselves.

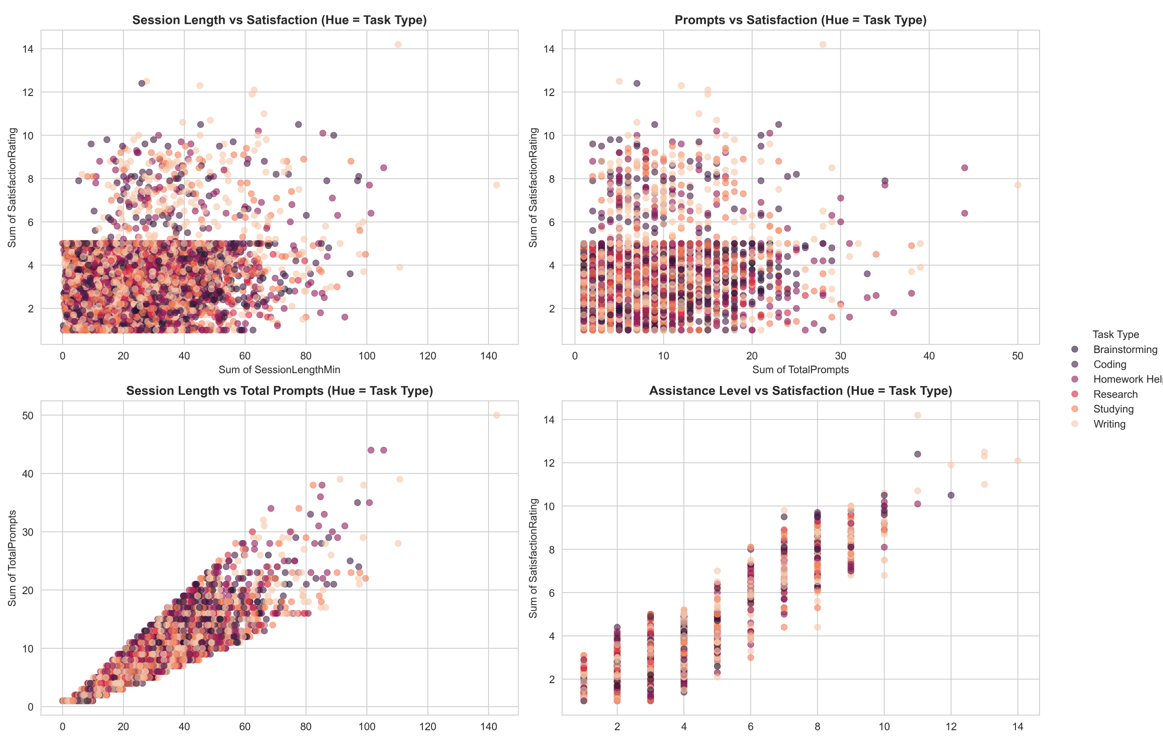

Figure 7: Multivariate Analysis by Task Type

There is no strong linear correlation between satisfaction and either session length or total prompts, suggesting that longer or more prompt-heavy sessions do not necessarily lead to higher satisfaction. However, there is a potentially strong positive correlation between satisfaction and assistance level, indicating that students who request more AI help tend to report higher satisfaction.

Session length and total prompts do exhibit a strong positive correlation, implying that longer sessions typically involve more prompts, which aligns with expectations about engagement depth. However, the variance doesn't seem to keep consistent at higher values, indicating that while longer sessions often have more prompts, this relationship is not strictly linear.

4.4 Focused Linear Correlations

We conducted a more in-depth examination of the identified linear relationships. The scatter grids from Figure-7 show only a couple of relationships that trend linearly. We therefore concentrate Pearson's R (Pearson [1909]) on the assistance-to-satisfaction and prompts-to-session-length pairings rather than forcing a full matrix.

Assistance and satisfaction continue to show a strong linear tie (coefficient of 0.84), supporting a potential regression fit. Session length and total prompts also correlate strongly (coefficient of 0.9), reinforcing the workload story without implying satisfaction gains directly, however given the variance increasing pattern observed above, a non-linear model might better capture that relationship.

Modeling the satisfaction–assistance relationship (Figure-8) yields an R_2 of 0.72 with strong statistical significance, meaning assistance alone explains most of the variance among the predictors we tested. Because the fitted points fall in a tight band, it suggests that increasing support intensity should reliably lift student satisfaction.

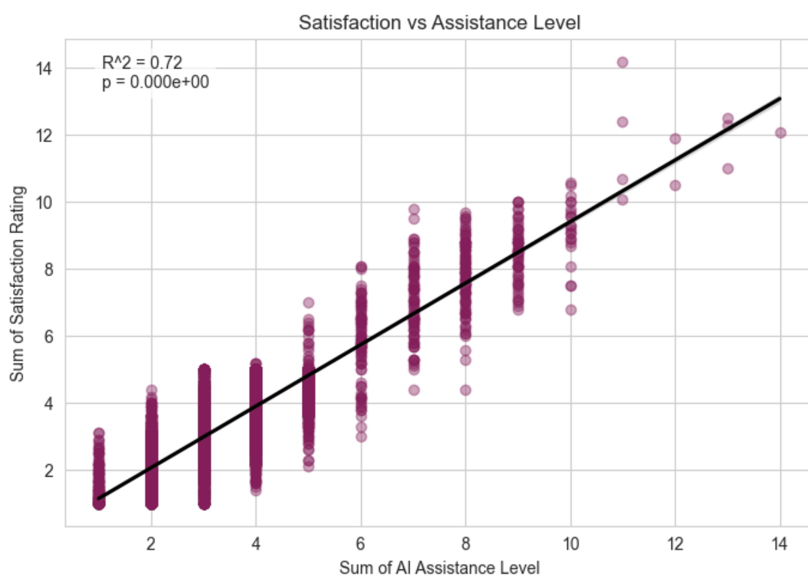

Figure 8: Satisfaction vs Assistance Level: Linear Regression

4.5 Temporal Dynamics

To identify seasonal spikes or drop-offs in AI reliance, we tracked both the volume of engagement and user satisfaction over time. We analyzed the temporal data by segmenting the year into quarters to obtain a clear overview of the annual trends.

With an even distribution of records across all quarters, the quarterly average evolution (Figure-9) reveals only negligible shifts in the mean. To formally assess the significance of these modest changes, we utilized one-way ANOVA tests (Fisher [1925]) to compare the means of satisfaction, session length, prompts, and assistance across the four quarters.

The quarter-level ANOVA (Table-4) (Fisher [1925]) indicates that average satisfaction, session length, prompt volume, and assistance intensity stay statistically flat across Q1–Q4 (p-values all >

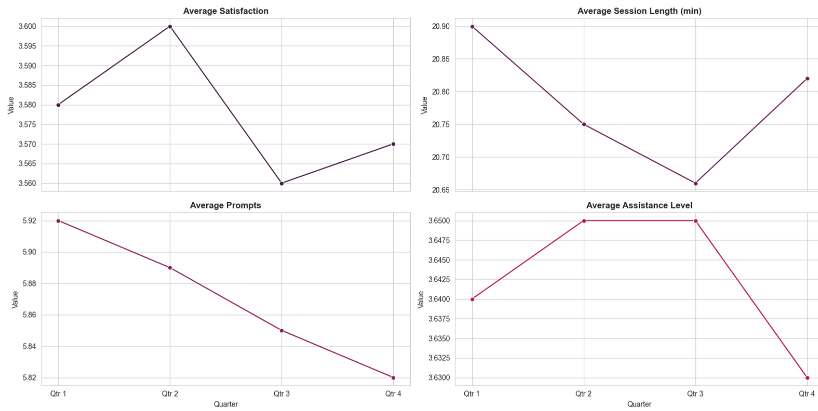

Figure 9: Quarter-Level Evolution

0.70). In practice, this means our learners behave similarly regardless of quarter: there is no evidence of seasonal spikes or slumps driving big swings in workload or satisfaction.

Table 4: One-Way ANOVA Test Results for Quarterly Averages

Metric	F-statistic	p-value	Significant?
Satisfaction Rating	0.44	0.7251	No
Session Length (min)	0.11	0.9534	No
Total Prompts	0.19	0.9066	No
Assistance Level	0.12	0.9511	No

4.6 Repeat Usage Behavior

Because reusing a service is important, we also focused on determining which students come back to the AI assistant and how satisfaction aligns with repeat adoption (Figure-10). Overall repeat usage sits at 69.7%, confirming that nearly seven in ten sessions come from return users.

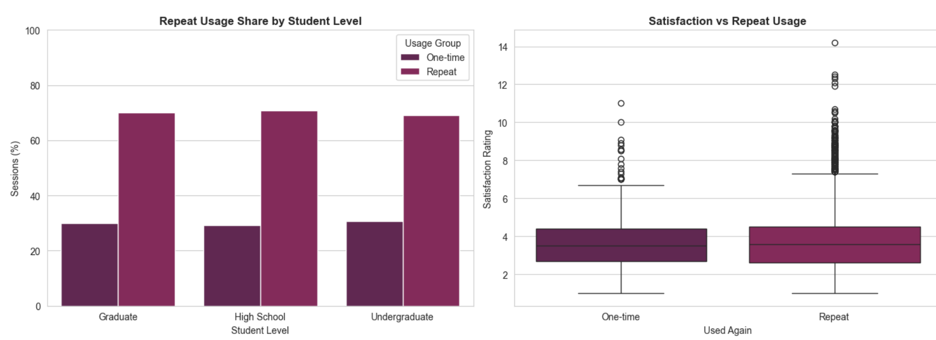

Figure 10: Repeat Usage Behaviour by Satisfaction and Student Level

The percentages of each subcategory of users can be observed in Table-5, Table-6 and Table-7. Satisfaction rises modestly for repeat users (mean 3.62 vs 3.49 for one-time users), reinforcing the small but significant lift observed in the Welch test we used check for statistical significance. The Welch t-test (Lu [2010]) showed the satisfaction gap between repeat and non-repeat users is statistically significant (p ¡ 0.001), but the effect size is small. In other words, repeat users score slightly higher on satisfaction on average, yet the practical lift is modest, so we should probe other drivers alongside repeat usage when prioritizing interventions.

Each student level holds a similar repeat share (70%), so differences in raw counts stem from enrollment volume rather than varying loyalty. Task types all clear the 60% repeat mark, with writing

Table 5: Summary Statistics for Satisfaction Rating by Re-usage Status

UsedAgainBool	Mean	Std Dev	Count
False	3.485413	1.198995	2893
True	3.618880	1.470382	6658

Table 6: Repeat Usage Share by Student Level (%)

Student Level	One-time (%)	Repeat (%)
Graduate	30.0	70.0
High School	29.1	70.9
Undergraduate	30.8	69.2

and coding contributing the largest returning cohorts, signalling where continuity programs may have the biggest reach. Given the limited satisfaction gap, deeper improvements will likely come from task- or segment-specific support rather than blanket repeat-use initiatives.

Table 7: Repeat Usage Counts by Task Type

Task Type	One-time	Repeat	All
Brainstorming	145	329	474
Coding	495	1370	1865
Homework Help	601	1271	1872
Research	170	305	475
Studying	640	1333	1973
Writing	842	2050	2892
All	**2893**	**6658**	**9551**

5 Discussion

Our exploratory analysis unfolded in several logical stages. We began with basic descriptive statistics and examinations of univariate distributions, then moved on to investigating bivariate relationships and conducting segmentation analyses, and finally examined temporal and behavioral patterns. Given that AI usage is a widely studied topic in the literature, we sought to benchmark our findings against those reported in comparable research.

Initial univariate analysis revealed that the "typical" student session is relatively short, with usage distributions heavily skewed toward quick interactions. Moving into bivariate analysis, we sought to explain what drives successful outcomes within these short windows. The distinct lack of correlation between session duration and satisfaction challenges the traditional assumption that more time spent on a task equates better learning outcomes. Instead, our findings align with preliminary reports by Zendy (zendy.io [2025]), where the data suggest that students treat AI not as a tutor that requires prolonged engagement, but as an efficiency tool. The value of the platform is defined by its ability to minimize, rather than extend, the time spent on academic labor.

A critical insight emerged during the segmentation phase regarding the undergraduate cohort. While this group represents the highest volume of consumers, they reported statistically lower satisfaction compared to High School and Graduate students. This inverse relationship between adoption and satisfaction suggests a "capability mismatch". A mismatch between user expectations and user experience can lead to dissatisfaction with and lower adoption rates of AI systems (Kinney [2024]).

Undergraduate coursework often sits in a difficult middle ground: it requires more nuance and critical rigor than high school work (which current AI models handle easily) but lacks the hyper-specialization of graduate work (where the user is often a mature expert capable of guiding the AI). This friction suggests that while undergraduates are eager to adopt these tools, the current "one-size-fits-all" models may not yet be tuned to the specific critical thinking standards required at the university level.

When analyzing performance by task type, our results highlight a tension between generative and analytical capabilities. Although "Writing" is the dominant use case, it yields lower satisfaction than "Research." This supports Adhikari's observation (Adhikari [2024]) that students are increasingly comfortable using AI as a synthesis engine for information retrieval, despite potential integrity concerns.

Finally, our temporal analysis revealed a flatness in seasonal trends that is statistically significant. The absence of "exam panic" spikes or mid-semester lulls, combined with the 70 percent retention rate identified in our behavioral analysis, suggests that AI usage has normalized. It is no longer a novelty reserved for emergencies but has become a piece of invisible infrastructure in the daily academic workflow, much like a search engine or word processor.

6 Conclusions

This comprehensive analysis of student-AI interaction data offers significant insight into usage patterns, satisfaction drivers, and adoption behaviors. By conducting our evaluation, we have observed several distinct patterns that have emerged that define the current state of student engagement with the AI tool.

The most powerful predictor of student satisfaction is the level of AI assistance requested. Students who engaged AI for higher levels of support consistently reported better outcomes, suggesting that the depth of the AI's contribution matters more to the students than the duration of their interaction. The strong relationship between assistance level and satisfaction—coupled with the lack of correlation with session length—indicates that students value efficiency and depth over prolonged interaction. Users measure value by how much they can rely on it to offload complex parts of their workload, rather than by the amount of time spent with the interface.

A clear dichotomy exists in the tool's utility. It currently serves as a robust engine for generative tasks like Writing, where satisfaction is seen to be maximized. However, the lower performance in information retrieval (Research) suggests that the current model is better optimized for content creation than for web inquiry or synthesis. This discrepancy highlights a critical area for future development if the tool is to become a holistic academic assistant. Technical improvements or user training focused on information retrieval could significantly boost overall value.

A key limitation posed in our analysis is that it relied exclusively on quantitative measures without accounting for qualitative context, which could explain more aspects of statistical significance. Incorporating qualitative feedback through surveys or interviews would provide richer context for the quantitative patterns observed. Longitudinal tracking of individual students would enable growth curve analyses and causal inference about skill development. Experimental designs comparing AI-assisted learning with traditional methods could establish the relative effectiveness of the tool. Finally, developing predictive models for at-risk students—those likely to discontinue usage or report low satisfaction—could enable proactive intervention and support optimization.

Ultimately, the stability of usage trends across the academic year, combined with a high rate of recurring users, points to a lucrative model to perfect. It demonstrates that AI assistance is no longer an experiment but an established component of student learning—one ready for its next phase of development.

References

S. Adhikari. Impact of ai in education processes, 2024. URL https://doi.org/10.18738/T8/RXUCHK. [Digital Poster].

D. Baariq. Data usage ai in school. Kaggle.com, 2025. URL https://www.kaggle.com/datasets/danishbaariq/data-usage-ai-in-school. Accessed 9 Dec. 2025.

B. Edwards. Why not robot teachers: Artificial intelligence for addressing teacher shortage. *Applied Artificial Intelligence*, 2018. doi: 10.1080/08839514.2018.1464286. URL https://doi.org/10.1080/08839514.2018.1464286.

Ronald A. Fisher. *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh, 1925.

H.J. Keselman. The tukey multiple comparison test: 1953–1976. *Psychological Bulletin*, 1977. doi: 10.1037/0033-2909.84.5.1050. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC3916511/.

M. Kinney. Expectation management in ai: A framework for understanding stakeholder trust and acceptance of artificial intelligence systems. *Heliyon*, 10(7):e28562, 2024. ISSN 2405-8440. doi: 10.1016/j.heliyon.2024.e28562.

Z. Lu. Welch's t test. *ResearchGate*, 2010. doi: 10.13140/RG.2.1.3057.9607. URL https://www.researchgate.net/publication/301292970_Welch.

K. Pearson. Determination of the coefficient of correlation. *Science*, 1909. doi: 10.1126/science.30.757.23. URL https://www.science.org/doi/10.1126/science.30.757.23.

X. Tan. Artificial intelligence in teaching and teacher professional development: A systematic review. *Computers and Education: Artificial Intelligence*, 2025. doi: 10.1016/j.caeai.2024.100355. URL https://doi.org/10.1016/j.caeai.2024.100355.

zendy.io. Ai in education for students and researchers: 2025 trends and statistics, 2025. URL https://zendy.io/blog/ai-in-research-for-students-researchers-2025-trends-statistics.