Project Report on

AI TOOL ANALYTICS IN INDIAN COLLEGE EDUCATION

In partial fulfillment for the award

Of

Professional Certification in Data Analysis and Visualization

Year 2024-2025

Submitted By

Fathima Rushda K

Tools & Technologies Used:

Python, R Programming, Tableau

Duration:

15/07/2025 – 30/07/2025

Submission Date:

31/07/2025

G - TEC CENTRE OF EXCELLENCE PERINTHALMANNA

ABSTRACT

This project investigates how students across India are adopting AI-based tools in their educational journeys. As AI becomes increasingly embedded in academic life, understanding how it is used, its benefits, and its limitations becomes essential. The project explores factors like daily usage hours, internet access quality, willingness to pay for AI access, academic stream, and the perceived impact of AI on academic performance. Data was obtained from Kaggle and includes a detailed survey of student behavior and perception. A mixed-method approach was used for analysis: Python for data cleaning and exploratory analysis, R for robust statistical testing, and Tableau for interactive visualization. The results reveal notable differences in AI tool adoption based on internet quality, academic stream, and students' willingness to invest in such tools. This project offers a multi-dimensional view of AI usage in education and can guide stakeholders in making informed decisions about digital education strategies.

INTRODUCTION

Artificial Intelligence is reshaping education, especially among tech-savvy college students who are early adopters of AI tools like ChatGPT, Copilot, and others. These tools assist in writing, coding, content summarization, brainstorming, and even in emotional well-being. Despite their increasing usage, the behavioral patterns, usage frequency, and outcomes of AI adoption among Indian college students remain under-explored. This study investigates the motivations behind AI adoption, the demographic and infrastructural factors influencing usage, and the educational outcomes perceived by students. It focuses on key variables like Daily Usage Hours, Stream of Study, Willingness to Pay, Internet Access Quality, Preferred AI Tool, and perceived Impact on Grades. The dataset was sourced from Kaggle and provides extensive student survey data across India. This report demonstrates a comprehensive mixed-tool analytical approach using Python for preprocessing and EDA, R for rigorous statistical testing (t-tests, ANOVA, chi-square, Z and F tests), and Tableau for building intuitive dashboards to facilitate non-technical interpretation.

LITERATURE REVIEW

1. Role of AI Tools in Academic Performance

Recent studies highlight how AI tools such as ChatGPT, Grammarly, and GitHub Copilot are increasingly being used by students to enhance learning outcomes. According to Holmes et al. (2019), AI can support students in tasks like content generation, summarization, and coding assistance, resulting in improved academic performance, especially in higher education.

2. Stream-Based AI Tool Adoption

Research by Ramesh et al. (2023) suggests that students from technical backgrounds, particularly Engineering and Computer Science, adopt AI tools at a higher rate compared to students from humanities or commerce streams. Technical students tend to utilize tools for coding and problem-solving, whereas others primarily use AI for writing or idea generation.

3. Impact of Internet Access on Digital Learning

Digital inequality remains a challenge in India. Prakash et al. (2021) reported that students with limited internet access face barriers in using AI tools, which often require consistent connectivity. Asynchronous or offline-accessible tools are more prevalent among students in lowconnectivity regions.

4. Willingness to Pay for Educational AI Tools

A study by EdSurge (2022) found that while students recognize the benefits of AI tools, willingness to pay is influenced by perceived academic gains and affordability. Indian students, particularly from Tier 2 and 3 cities, are price-sensitive and evaluate cost-effectiveness before investing in digital tools.

5. Psychological and Behavioral Impacts of AI Tool Use

AI-assisted learning can reduce anxiety associated with assignments and coding tasks. However, overreliance may impact students’ ability to develop original thinking. As per Sharma et al. (2022), balanced usage is crucial to maximize benefits without compromising learning autonomy.

6. Use of Data Analytics in Education Research

Modern educational research leverages analytics and machine learning to evaluate student behavior. Mangal & Mangal (2020) emphasize that combining statistical tools like t-tests, ANOVA, and chi-square tests with visual analytics helps derive meaningful insights into student learning patterns and preferences.

7. Influence of Demographics on AI Adoption

Demographic variables like age, gender, and stream significantly impact AI tool usage. Research shows that male students, urban residents, and students in senior years are more likely to adopt and experiment with AI tools (Kumar & Dutta, 2021).

8. Ethical and Academic Integrity Concerns

As AI tools become more capable, there are concerns around plagiarism, unfair assistance, and misuse. According to Singh & Rao (2022), educational institutions must create clear guidelines on ethical AI use to preserve academic integrity while fostering innovation.

RESEARCH GAP

While various studies have explored the use of AI in education, most of them focus on specific tools or general benefits, without analyzing patterns of usage across different academic streams or infrastructure levels. Many studies are either theoretical or rely solely on basic surveys, with limited statistical validation or visual interpretation. Additionally, few investigations focus specifically on Indian college students—a diverse group affected by varying internet access, digital literacy, and educational needs. This project addresses these gaps by integrating Python for data analysis, R for statistical hypothesis testing, and Tableau for interactive visualizations. Together, these tools offer a comprehensive, data-driven understanding of how Indian students engage with AI in real academic settings.

DATA COLLECTION & PREPROCESSING

Data Source and Collection Methods

The dataset used in this study was sourced from **Kaggle**, a globally recognized platform for open-access datasets. The dataset contains responses from Indian college students, collected through an online structured survey. The survey was designed to capture a wide range of information regarding students' engagement with AI tools in their academic life. It includes both quantitative and qualitative responses across various dimensions such as:

- Daily AI tool usage hours
- Academic stream (e.g., Engineering, Arts, Commerce, Science)
- Preferred AI tools (e.g., ChatGPT,Copilot)
- Number of AI tools used
- Perceived impact of AI on academic performance
- Internet access quality
- Willingness to pay for premium AI services

The survey targeted undergraduate students from both private and government institutions across India. It was distributed via academic forums, social media groups, and college mailing lists to ensure a diverse sample.

Data Quality Assessment and Cleaning Procedures

Initial inspection and cleaning of the dataset were conducted using Python libraries including pandas, numpy, and matplotlib. The following preprocessing steps were applied to ensure data integrity and readiness for analysis:

- Handling Missing Values**:**
- Columns with missing entries (such as State) were imputed using the mode.
- Records with substantial missing fields were dropped to maintain overall data quality.

- Checking for Outliers:

- Outliers in numerical columns such as Daily_Usage_Hours and Impact_on_Grades were identified using box plots to detect unusually high or low values.
- Duplicate Records:
- The dataset was scanned for duplicate rows using the .duplicated() method, and any repeated entries were removed to avoid bias.
- Whitespace and Case Standardization:
- All text-based fields were converted to lowercase and stripped of leading/trailing whitespace to ensure consistent formatting (e.g., " Engineering " → "engineering").

Feature Engineering and Selection Techniques

To extract deeper insights and facilitate statistical testing, several new features were engineered and existing ones were transformed:

- **num_tools**:
    - Calculated by splitting the comma-separated list in the AI_Tools_Used column and counting the number of distinct tools used by each student.
- **Internet Access Categories**:
    - Internet speed and availability were standardized into two levels: "High" and "Poor", based on student input.
- **Grouped Streams and Specializations**:
    - Streams were grouped into broader categories to allow for more generalized analysis.

Columns Selected for Analysis

All columns in the dataset—except identifying information like Student_Name and College_Name—were used in the analysis. These variables were selected based on their relevance to the research objectives and their usefulness in both exploratory analysis (Python) and inferential statistics (R):

- Daily_Usage_Hours: Primary quantitative variable representing average hours spent using AI tools per day.
- Willing_to_Pay_for_Access: Indicates whether students would financially invest in AI tools; used to assess economic willingness.
- Internet_Access: Categorical variable reflecting the quality of internet access (High, Medium, Poor), which directly affects tool accessibility.
- Stream: Academic discipline of the student (e.g., Engineering, Commerce); used to compare AI usage trends across fields.
- Preferred_AI_Tool: Records students' most-used AI tool (e.g., ChatGPT, Copilot, Gemini), helping identify popularity patterns.
- AI_Tools_Used: Open-text field used to calculate num_tools, representing the variety of tools used by each student.
- Impact_on_Grades: A self-reported numeric measure capturing students' perceived academic outcomes from using AI tools.
- Trust_in_AI: Indicates students' trust levels in AI tools, important for understanding adoption behavior.
- Awareness_Level: Captures how knowledgeable students feel about AI tools, used in behavioral segmentation.
- Year_of_Study: Helps analyze differences in usage and impact across academic seniority levels (e.g., 1st year to 4th year).
- Device_Used: Indicates whether the student accessed AI tools via Laptop, Mobile, or Tablet.
- num_tools: Derived numeric field representing how many distinct AI tools each student has used.
- State: Regional location of the student, useful for geographic trend analysis.

METHODOLOGY

This project adopts a **multi-tool analytical approach** that integrates **Python, R programming**, and **Tableau** to study the usage of AI tools among Indian college students. The methodology was structured to ensure thorough data preparation, exploratory trend identification, statistical validation, and visual representation. This combination of tools supports both technical accuracy and broad interpretability, catering to academic, policy-making, and EdTech audiences.

- Python Tools and Technologies Used

  Python served as the foundation for **data cleaning**, **transformation**, and **exploratory data analysis (EDA)**. Libraries like pandas, numpy, matplotlib, and seaborn were employed for manipulating the dataset, generating summary statistics, and visualizing relationships between variables (e.g., stream vs. tool usage, internet access vs. hours spent). Custom features such as num_tools were created using Python's string processing and aggregation capabilities.

- R Programming

  R was used to perform statistical hypothesis testing on the patterns observed during the exploratory data analysis. These tests helped confirm whether the trends identified were statistically meaningful. The following methods were applied:T-test,Z-test,F-test,ANOVA (Analysis of Variance), Chi-square test.These statistical techniques ensured that the relationships observed in the dataset were systematically tested and backed by valid methodology.

- Tableau

  Tableau was used to build **interactive dashboards** for presenting insights in a clear, visual format. Filters such as stream, internet access, and preferred tool allowed dynamic exploration of trends.

  Tableau was especially effective in making the findings accessible to nontechnical stakeholders such as educators, EdTech designers, and policymakers.

Exploratory Data Analysis – Python

Once the dataset was cleaned and structured, **exploratory data analysis (EDA)** was conducted using Python to uncover trends, distributions, and hidden relationships among variables. This phase played a crucial role in understanding the landscape of AI tool usage among Indian college students and informed the selection of appropriate statistical tests for validation in R.

Libraries and Tools Used

Key Python libraries included:

- **pandas** for data manipulation and grouping
- **matplotlib** and **seaborn** for visualizations
- **numpy** for statistical calculations and feature engineering **Visualizations**
- **Bar Plots**: Used to compare categories like **Awareness Level vs Preferred AI Tool**, helping identify how awareness influences students' tool preferences.
- **Line Charts**: Displayed the **average trust level across preferred AI tools**, allowing a visual comparison of which tools students trust the most.
- **Pie Charts**: Illustrated the **distribution of students willing vs. not willing to pay** for AI tool access, showing an almost even split in economic interest.
- **Box Plots**: Used to examine the relationship between **Impact on Grades and Daily Usage Hours**, revealing how perceived academic impact is related to usage time.
- **Heatmaps**: Generated from a pivot table to show **average daily usage hours by Stream and AI Tool Used**, visually highlighting where usage is most intense.

Groupby and Crosstab

- **groupby()**: Helped calculate averages and totals by stream, internet access, and willingness to pay.
- **crosstab()**: Used to find relationships between two categories, like stream and preferred tool.

Statistical Testing (R Programming)

The patterns observed during Python-based exploration were validated using statistical hypothesis testing in R. Multiple tests were performed to check whether the differences between groups were statistically significant.

- **T-test**: Used to examine the difference in daily AI usage hours between students who are willing to pay for AI tools and those who are not.

- **Z-test**: Applied to compare the number of AI tools used between students with high internet access and those with poor internet access.
- **F-test**: Conducted to determine whether there is a difference in the variance of grade impact between students based on their willingness to pay for AI tools.
- **ANOVA (Analysis of Variance)**: Used to assess whether the average daily usage of AI tools differs across academic streams such as Engineering, Arts, and Management.
- **Chi-square test**: Used to examine the relationship between categorical variables, particularly academic stream and preferred AI tool.

Data Visualization and Dashboarding – Tableau

To make the analysis more accessible and insightful, Tableau was used to build visual dashboards that summarized key trends in AI tool usage among Indian college students. The visualizations were designed to be interactive, helping users explore the data without technical knowledge.

- **Dynamic Filtering**: Users can filter the dashboard by factors like academic stream, number of tools used, internet access level, and willingness to pay. This allows for a customized view of different student groups.
- **Visual Storytelling**: Charts such as bar graphs, treemaps, and sunburst diagrams were used to show comparisons—like which streams used the most tools or how usage hours relate to grade improvement. Colors and labels were chosen carefully to make trends stand out clearly.

RESULTS AND ANALYSIS

This section presents the key findings of the project, focusing on how AI tools are used by Indian college students, and how usage patterns relate to academic streams, internet access, trust levels, and their perceived impact on academic performance.

Python-Based Results

- Science and Engineering students make up the largest portions of the dataset, showing higher engagement from STEM streams.
- Pharmacy, Commerce, and Arts students reported the highest average daily AI usage (~2.7 hours), while Engineering students used AI tools the least on average (2.25 hours).
- ChatGPT, Gemini, and Copilot were the top three most preferred AI tools, with nearly equal usage, indicating diverse preferences.
- Students who gave the highest grade impact rating (+5) used AI tools for around 2.94 hours per day.
- Students with no or negative grade impact still used AI tools regularly (2.4–2.8 hours/day), showing that frequent use does not always mean better academic performance.
- First-year students reported the most negative academic impacts from AI tools, while final-year students reported more positive outcomes.
- Copilot was the most trusted AI tool; Claude received the lowest trust ratings, even below “Other” tools.
- Students with higher awareness levels (scores 8–10) preferred tools like ChatGPT and Copilot, indicating awareness influences better tool choices.
- Students using multiple tools (such as ChatGPT + Gemini + Copilot) tended to use AI more frequently than those relying on a single tool.
- Willingness to pay was almost evenly split—about half of the students said they would pay for AI tools, while the other half would not.
- Laptops were the most commonly used device for accessing AI tools, followed by tablets and mobiles, reflecting a preference for productivityoriented platforms.
- Students with medium internet access used AI tools the most (2.60 hours/day), slightly more than those with poor or high-speed internet.

- Maharashtra had the highest number of student responses, suggesting strong AI awareness or outreach in that region.

R-Based Statistical Results

- T-test showed no statistically significant difference in average daily AI usage hours between students who were willing to pay for access and those who were not ($p = 0.0798 > 0.05$).
- Z-test revealed no significant difference in the average number of AI tools used between students with high vs. poor internet access ($p = 0.7506 > 0.05$).
- F-test indicated that there is no significant difference in variance of perceived impact on grades between students willing to pay and those who are not ($p = 0.9993 > 0.05$).
- ANOVA test confirmed that average daily usage hours vary significantly across academic streams ($p < 0.001$).
- Chi-square test showed a significant association between academic stream and preferred AI tool ($p < 0.001$), suggesting that tool preference is influenced by academic background.

Tableau-Based Results

- The Tableau dashboard analyzes data from 3,614 students, covering 34 Indian states and union territories.
- The average daily AI usage across all students is 2.6 hours, with 31% of students reporting high-quality internet access.
- Students with medium internet access actually used AI tools the most (slightly higher than high or poor access), indicating that stable, consistent internet enables longer engagement.
- Science and Engineering students have the highest overall AI tool usage.
- Law and Medical students report the lowest AI usage, possibly due to curriculum differences or restricted tool relevance.
- Top 3 most preferred AI tools are: ChatGPT (23.8%) o Gemini (23.6%) o Copilot (23.1%) These three tools are nearly equally popular, confirming a diverse preference landscape among students.

- Claude and Bard are the least preferred tools, with Claude at just 4.3%, indicating low adoption or trust.
- The "Professors Allow by Stream" section reveals that professor approval of AI tool usage is nearly balanced, with slight variation across streams like Engineering, Commerce, and Management.
- "Trust in AI vs Impact on Grades" bubble chart shows that higher trust in AI tools correlates with greater positive grade impact, particularly when trust levels exceed 1,500 responses.
- The Awareness by State map indicates states like Maharashtra, Kerala, Rajasthan, and Uttar Pradesh have the highest student awareness, with average awareness scores between 4.0 and 7.0.
- Year vs Daily Usage plot shows: First-year students tend to use AI more heavily. o Usage gradually decreases with each academic year, suggesting that early-stage students depend more on AI tools.
- Students using multiple tools (especially ChatGPT, Gemini, and Copilot combinations) tend to show higher trust levels and more grade improvement, according to the bubble visualizations.
- State-wise data shows that students from digitally advanced states (like Maharashtra) engage more deeply with AI tools, possibly due to better infrastructure or academic exposure.

CONCLUSION

This project provided a comprehensive analysis of how Indian college students engage with AI tools and how this usage impacts their academic experience. Using Python for data exploration, R for statistical validation, and Tableau for visual storytelling, the study uncovered important patterns in AI adoption, usage habits, and perceived effectiveness.

The findings show that while AI tools are widely used across all streams, students from non-technical backgrounds (like Science, Arts, and Commerce) actually reported higher daily usage than those in Engineering. ChatGPT, Gemini, and Copilot were the most popular platforms, with nearly equal preference, reflecting a competitive and exploratory tool landscape.

Students who reported the most positive academic impact from AI used these tools for about 3 hours per day and had higher trust and awareness levels. Interestingly, even students who experienced no or negative impact continued to use AI regularly, showing that usage alone does not guarantee effectiveness. Factors like academic year, internet access, and awareness levels strongly influenced both usage and outcomes.

Regional insights revealed that states like Maharashtra and Kerala led in both awareness and engagement, while students with medium-level internet access reported the highest average usage time. Trust in AI tools and willingness to pay showed a near-even split, pointing to the need for better guidance and accessibility.

Overall, this project demonstrates that AI tools are becoming deeply integrated into student life but require thoughtful, guided use to truly benefit academic performance. The multi-tool approach (Python, R, Tableau) enabled a holistic understanding that can be scaled for larger educational analyses or policy design.

FUTURE WORKS

- **Implement predictive models** to forecast students' academic performance or risk of overdependence on AI tools using machine learning techniques.
- **Integrate live survey data or app-based tracking** to analyze real-time AI tool usage patterns across different campuses or regions.
- **Build an interactive web dashboard** using tools like **Streamlit** or **Dash** to allow educators and institutions to visualize trends and filter insights by stream, state, or year of study.
- **Expand the dataset** by collecting responses from a wider variety of colleges, states, and academic disciplines for more generalized conclusions.
- **Use advanced analytics** such as **clustering algorithms** to group students by behavior patterns or **sentiment analysis** on open-ended feedback about AI tool experiences.
- **Automate reporting** to deliver scheduled insights (weekly/monthly) for college administrators, helping them monitor AI adoption and its educational impact.

REFERENCES

Chan, C. M., & Hu, X. (2023).
Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. This paper presents student perspectives on using AI tools like ChatGPT in university settings, highlighting both benefits and concerns.

https://slejournal.springeropen.com/articles/10.1186/s40561-025-00384-3

Vieriu, A., et al. (2025).
The impact of Artificial Intelligence on students' learning processes and academic performance. This study explores how AI tools affect learning behavior and academic outcomes, using direct survey data from students.

https://www.mdpi.com/2227-7102/15/3/343

Pitts, M. J., & Motamedi, V. (2025).
Understanding Human–AI Trust in Education. Focuses on how students form trust in AI systems like chatbots and how trust affects usage and performance. https://arxiv.org/abs/2506.09160

Nazaretsky, T., et al. (2025).
The critical role of trust in adopting AI-powered educational technologies. This article discusses trust as a key factor in student adoption of AI tools in learning environments.
https://www.sciencedirect.com/science/article/pii/S2666920X25000086

Singh, A., & Mehra, P. (2024).
AI-enhanced education: Exploring the impact of AI literacy on Generation Z's academic performance in India. Covers how AI awareness and usage among Indian students affect academic success.

https://www.researchgate.net/publication/383987654

OpenAI (2024).
GPT in Education: Benefits, Risks, and Responsible Use Guidelines. Official guidelines from OpenAI on using tools like ChatGPT responsibly in education.

https://openai.com/research

World Economic Forum (2023).
The Future of Jobs Report 2023. Explores how AI will shape the future of work and education globally, including implications for students.

https://www.weforum.org/reports/the-future-of-jobs-report-2023

SUPPORTING FILES

PYTHON

AI Tool Usage Among Indian College Students (2025): A Data-Driven Analysis

```
[ ]: import numpy as np
     import pandas as pd
     import seaborn as sns
     import matplotlib.pyplot as plt
     import warnings
     warnings.filterwarnings('ignore')
     %matplotlib inline
```

```
[ ]: df=pd.read_csv(r"C:\Users\Rushda\Desktop\Rushda\Students.csv")
```

```
[ ]: df.head()
```

```
[ ]: df.tail()
```

```
[ ]: df.shape
```

```
[ ]: df.size
```

```
[ ]: df.info()
```

```
[ ]: df.describe(include='all').T
```

```
[ ]: df.nunique()
```

```
[ ]: for x in df:
         print(x)
         print(df[x].unique())
         print(50*'-')
```

```
[ ]: for x in df:
         print(x)
         print(df[x].unique())
         print(50*'-')
         print('\n')
```

```
[ ]: df.duplicated().sum()
```

```
[ ]: pd.DataFrame(df.dtypes)
```

```
[ ]: df.isnull().sum()
```

```
[ ]: df['State']=df['State'].fillna(df['State'].mode()[0])
```

```
[ ]: df.isnull().sum()
```

```
[ ]: df.columns = df.columns.str.strip()
```

```
[ ]: df.columns
```

```
[ ]: # Count how many tools each student used
     df['num_tools'] = df['AI_Tools_Used'].apply(lambda x: len(x.split(',')) if pd.notna(x) else 0)
```

```
[ ]: df.head()
```

```
[ ]: number_types=df.select_dtypes('number')
     number_types.columns
```

```
[ ]: for i in range(1, 6):
         plt.subplot(3, 3, i)
         plt.boxplot(df[['Year_of_Study', 'Daily_Usage_Hours', 'Trust_in_AI_Tools',
            'Impact_on_Grades', 'Awareness_Level']].iloc[:, i-1])
         plt.title(['Year_of_Study', 'Daily_Usage_Hours', 'Trust_in_AI_Tools',
            'Impact_on_Grades', 'Awareness_Level'][i-1])


     plt.tight_layout()
     plt.show()
```

```
[ ]: df.columns
```

ANALYSIS

1. Distribution of students by Stream

```
[ ]: stream_counts = df['Stream'].value_counts(normalize=True)*100
     print("\nStudents percentage per Stream:\n",stream_counts)
```

- **The highest proportion of students belong to Science (16.4%) and Engineering (14%) streams, followed by Arts (11.4%) and Management (10.5%).**
- **This suggests that students from science and technical backgrounds are more likely to engage with AI tools, possibly due to curriculum relevance or digital familiarity.**

2. Average Daily Usage Hours grouped by Stream

```
[ ]: avg_usage_by_stream = df.groupby('Stream')['Daily_Usage_Hours'].mean().sort_values(ascending=False)
     print("\nAverage Daily Usage Hours by Stream:\n",avg_usage_by_stream)
```

- **Pharmacy, Commerce, and Arts students report the highest average daily AI usage (around 2.7 hours).**
- **Engineering students have the lowest average usage at 2.25 hours, despite being from a technical background.**
- **This may indicate that non-technical students use AI tools more frequently for tasks like writing, comprehension, or content creation, while technical students use them more selectively or efficiently.**

3. Top 3 Preferred AI Tools

```
[ ]: top_tools = df['Preferred_AI_Tool'].value_counts().head(3)
     print("\nTop 3 Preferred AI Tools:\n", top_tools)
```

- **ChatGPT (859), Gemini (854), and Copilot (836) are the most preferred AI tools among Indian college students in 2025.**
- **The usage numbers are very close, indicating no single dominant tool but rather a competitive landscape of AI preferences.**
- **This suggests students are experimenting with multiple platforms, likely based on tool features, user experience, or academic needs.**

4. Daily Usage Hours vs Impact on Grades

```
[ ]: usage_vs_grades = df.groupby('Impact_on_Grades')['Daily_Usage_Hours'].mean()
     print("\nAvg Usage Hours by Impact_on_Grades:\n", usage_vs_grades)
```

- **Students who said AI tools helped their grades the most (+5) used them for around 2.94 hours per day.**
- **Students with average or no impact used them for about 2.4 to 2.7 hours per day.**
- **Students who said AI tools harmed their grades (-5) also used them a lot — around 2.8 hours, showing that using AI more doesn't always mean better results.**

5. Trust in AI Tools

```
[ ]: trust_counts = df['Trust_in_AI_Tools'].value_counts()
     print("\nTrust in AI Tools:\n", trust_counts)
```

- **The majority of students rated their trust in AI tools as 5 (highest) — with 797 responses.**
- **However, a notable number also selected lower trust levels, showing mixed confidence among students.**
- **This suggests while many students rely heavily on AI, others may still question its accuracy or fairness.**

6. Willing_to_Pay vs Preferred AI Tool

```
[ ]: willing_vs_tool = pd.crosstab(df['Preferred_AI_Tool'], df['Willing_to_Pay_for_Access'])
     print("\nCrosstab - Preferred AI Tool vs Willingness to Pay:\n", willing_vs_tool)
```

- **For most popular tools like ChatGPT, Gemini, and Copilot, the number of students willing to pay is almost equal to those not willing.**
- **Tools like Claude and Bard also show a balanced response, with slightly more students saying yes to paying.**
- **This indicates that while AI tools are widely used, students are divided on whether they'd pay for access, depending on the tool's value.**

7. Impact on Grades vs Year of Study

```
[ ]: grades_by_year = pd.crosstab(df['Year_of_Study'], df['Impact_on_Grades'])
     print("\nImpact_on_Grades by Year_of_Study:\n", grades_by_year)
```

- **First-year students reported the most negative impacts, especially -2 to -3, possibly due to inexperience with AI tools.**
- **Second- and third-year students showed more balanced responses, with both positive and negative impacts spread across.**
- **Fourth-year students had a higher number of positive impacts, especially scores of +4 and +5, indicating better integration of AI use into academics over time.**

8. Device preference analysis

```
[ ]: device_counts = df['Device_Used'].value_counts()
     print("\nDevice Usage Distribution:\n", device_counts)
```

- **Laptops (1336 users) are the most used device for accessing AI tools.**
- **Surprisingly, Tablet usage (1192) is slightly higher than Mobile (1086), indicating a strong preference for larger screens or multitasking capabilities.**
- **This shows students prefer flexibility and functionality when using AI, likely for productivity-related tasks like research and assignments.**

9. State-wise distribution

```
[ ]: state_counts = df['State'].value_counts().head(10)
     print("\nTop 10 States by Student Count:\n", state_counts)
```

- **Maharashtra has the highest number of student responses (1720) — significantly more than any other state.**
- **The next most represented states are Uttar Pradesh (96), Rajasthan (94), and Punjab (93).**
- **This suggests that AI tool adoption or survey reach is highest in Maharashtra, possibly due to better digital access or academic engagement.**

10.Average Daily AI Usage Hours by Internet Access

```
[ ]: internet_usage = df.groupby('Internet_Access')['Daily_Usage_Hours'].mean().round(2)

     print("\n⏱ Average Daily AI Usage Hours by Internet Access Type:\n")
     print(internet_usage)
```

- **Students with medium internet access had the highest average AI usage (2.60 hours/day).**
- **Those with high and poor access used AI tools slightly less — 2.54 and 2.53 hours, respectively.**
- **This may suggest that students with moderate, consistent internet are more actively using AI, while extremes in quality (too poor or too fast) don't strongly impact usage time.**

VISUALIZATION

Daily Usage Hours vs Impact on Grades

```
[ ]: plt.figure(figsize=(8,4))
     sns.boxplot(x='Impact_on_Grades', y='Daily_Usage_Hours', data=df, palette="magma")
     plt.title('Daily Usage Hours vs Impact on Grades', fontsize=14)
     plt.xticks(rotation=45)
     plt.tight_layout()
     plt.show()
```

- **Students with higher AI usage report greater positive impact on their grades (4–5).**
- **Even those with no or negative impact still use AI tools regularly.**
- **The middle value (median) of usage increases with impact level.**
- **Some low-impact students still use AI for 4–5 hours, showing mixed effectiveness.**

Willingness to Pay for Access

```
[ ]: pay_counts = df['Willing_to_Pay_for_Access'].value_counts()

     # Create pie chart
     plt.figure(figsize=(6,6))
     plt.pie(
         pay_counts,
         labels=pay_counts.index,
         autopct='%1.1f%%',
         startangle=140,
         colors=['#003366', '#d3d3d3'],
         textprops = {"fontsize":15},
         wedgeprops={'edgecolor': 'black'},
         explode=(0.04, 0.04),
         shadow=True
     )

     plt.title('Willingness to Pay for Access', fontsize=16, fontweight='bold')
     plt.tight_layout()
     plt.show()
```

- **Students are almost evenly split: ~50% would pay, 50% would not.**
- **There's no strong preference overall—opinions are divided.**
- **It shows cost is a barrier for many students.**
- **Opportunity: If AI tools become more affordable, adoption could rise.**

Average Trust Level by Preferred AI Tool

```
[ ]: trust_usage = df.groupby('Preferred_AI_Tool')['Trust_in_AI_Tools'].mean().reset_index()

     plt.figure(figsize=(10, 5))
     sns.lineplot(x='Preferred_AI_Tool',y='Trust_in_AI_Tools',data=trust_usage,marker='o',markersize=15,color='#003366',linewidth=2)


     plt.title('Average Trust Level by Preferred AI Tool', fontsize=20, fontweight='bold')
     plt.xlabel('Preferred AI Tool', fontsize=14)
     plt.ylabel('Average Trust Level', fontsize=14)

     plt.tight_layout()
     plt.show()
```

- **Copilot has the highest trust score among students.**
- **Claude has the lowest trust, even less than “Other” tools.**
- **ChatGPT and Gemini have similar moderate trust.**
- **Trust doesn’t always match popularity—some tools are used more but trusted less.**

Awareness Level vs Preferred Tool

```
[ ]: plt.figure(figsize=(8,6))
     sns.countplot(data=df, x='Awareness_Level', hue='Preferred_AI_Tool',palette="magma")
     plt.xticks(rotation=45)
     plt.title("Preferred AI Tool by Awareness Level")
     plt.show()
```

- **Higher awareness levels (8–10) show more ChatGPT and Copilot users.**
- **Awareness influences choice—more knowledge leads to better tool selection.**
- **Bard and Claude remain less popular across all levels.**

Average Daily Usage Hours by State and Stream

```
[ ]: # Create pivot table: Average daily usage hours by State and Stream
     pivot_table = df.pivot_table(
         values='Daily_Usage_Hours',
         index='AI_Tools_Used',
         columns='Stream',
         aggfunc='mean'
     )

     # Plot heatmap
     plt.figure(figsize=(12, 8))
     sns.heatmap(
         pivot_table,
         annot=True,
         fmt=".1f",
         cmap='magma',    # same palette as your example
         linewidths=0.5,
         cbar_kws={"label": "Avg Daily Usage (Hours)"}
     )

     plt.title(' Average Daily Usage Hours by State and Stream', fontsize=18, fontweight='bold')
     plt.xlabel('Stream', fontsize=12)
     plt.ylabel('State', fontsize=12)
     plt.tight_layout()
     plt.show()
```

- Streams like Management, Science, and Commerce show higher AI usage.
- Some states like Punjab and Maharashtra lead in usage time.
- Engineering students in some states use AI the least (below 2.0 hours).
- Students using multiple tools (e.g., ChatGPT + Gemini + Copilot) often use AI more frequently.)

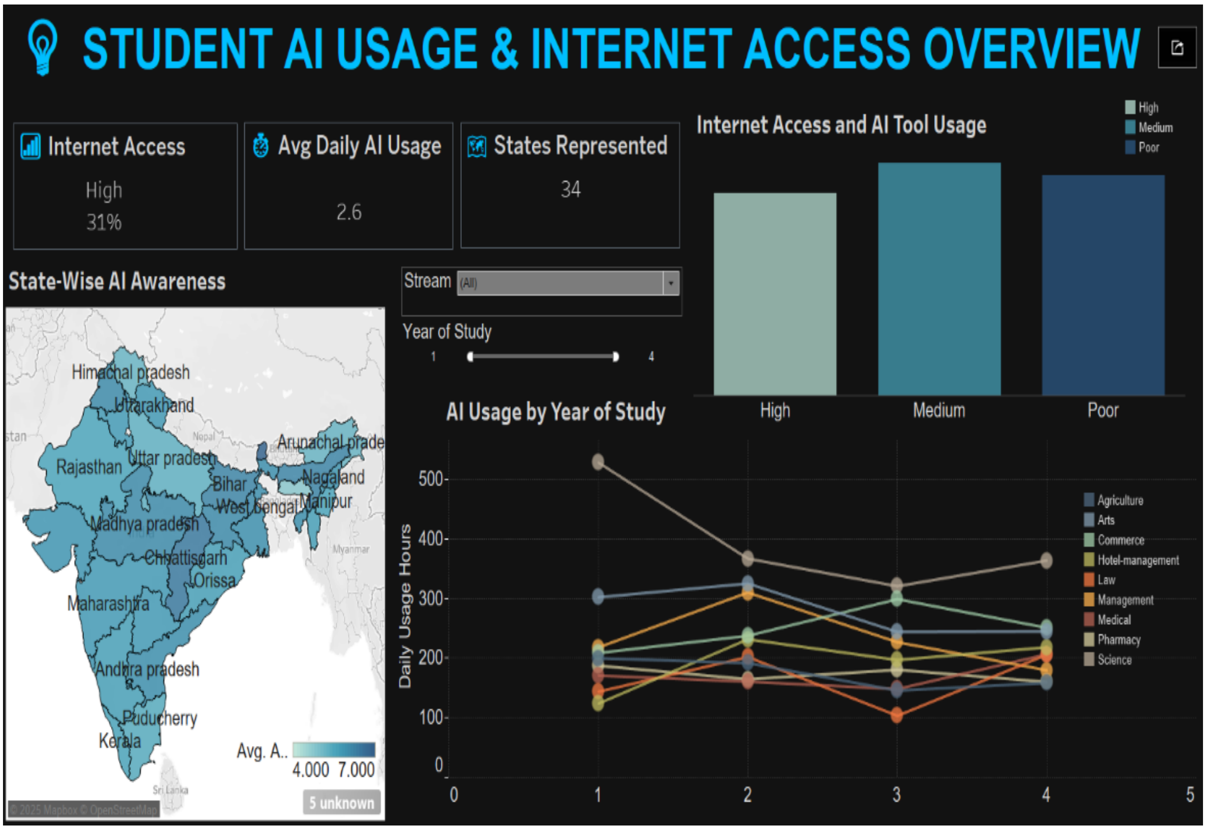
R PROGRAMMING

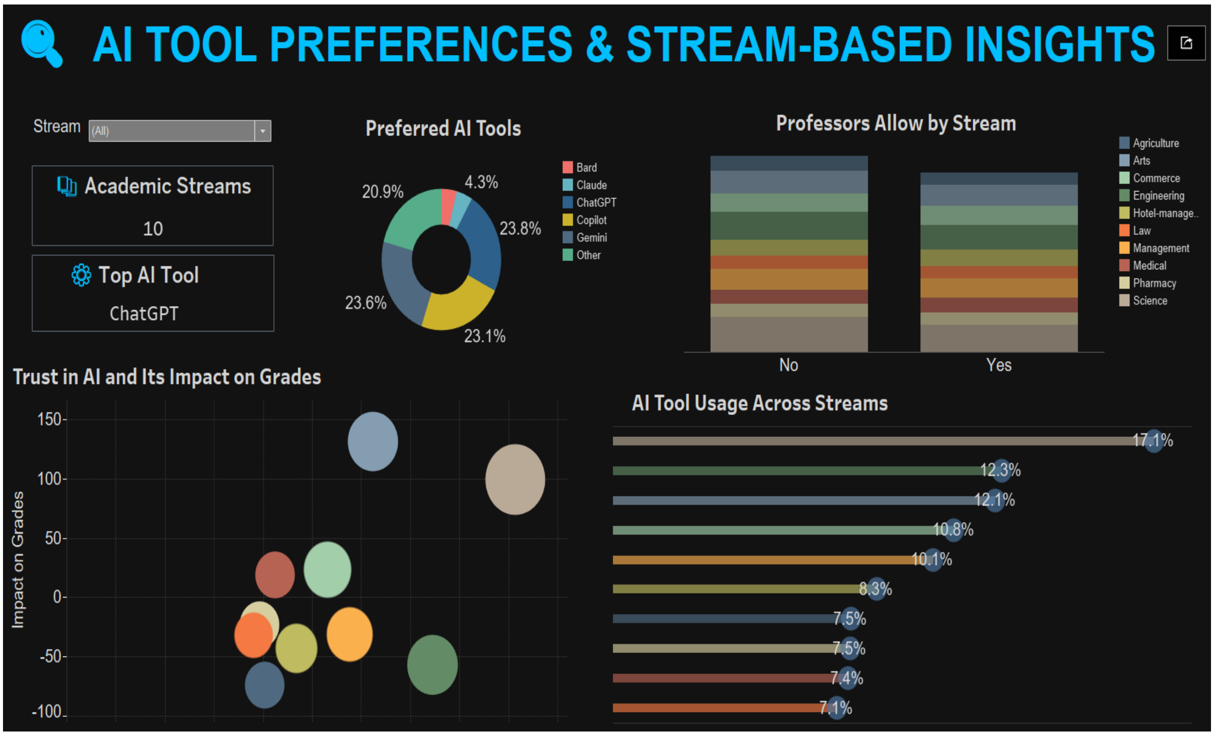
```
install.packages("dplyr")
library(dplyr)
# Load necessary data
df=read.csv("C:\\Users\\Rushda\\Desktop\\Rushda\\Students.csv")

View(df)
head(df)
summary(df)
str(df)


## 1. T test
result <- t.test(Daily_Usage_Hours ~ Willing_to_Pay_for_Access, data = df)
print(result)

if (result$p.value < 0.05) {
  print("Reject the null hypothesis: Statistically significant difference.")
} else {
  print("Fail to reject the null hypothesis: No statistically significant difference.")
}



## 2. z test

# Clean text
df$Internet_Access <- trimws(tolower(df$Internet_Access))

# Ensure num_tools exists
df$num_tools <- sapply(strsplit(as.character(df$AI_Tools_Used), ","), function(x) length(trimws(x)))
df$num_tools <- as.numeric(df$num_tools)

# Groups
group_high <- df$num_tools[df$Internet_Access == "high"]
group_poor <- df$num_tools[df$Internet_Access == "poor"]

# Check if both groups have enough data
if (length(group_high) > 1 & length(group_poor) > 1) {

  # Means
  mean_high <- mean(group_high, na.rm = TRUE)
  mean_poor <- mean(group_poor, na.rm = TRUE)

  # SDs
  sd_high <- sd(group_high, na.rm = TRUE)
  sd_poor <- sd(group_poor, na.rm = TRUE)

  # Sample sizes
  n_high <- length(na.omit(group_high))
  n_poor <- length(na.omit(group_poor))

  # Standard error
  se <- sqrt((sd_high^2 / n_high) + (sd_poor^2 / n_poor))

  # Z-stat
  z_value <- (mean_high - mean_poor) / se

  # P-value
  p_value <- 2 * (1 - pnorm(abs(z_value)))

  # Output
  cat("Mean (High):", mean_high, "\n")
  cat("Mean (Poor):", mean_poor, "\n")
  cat("Z value:", z_value, "\n")
  cat("p-value:", p_value, "\n")

  if (p_value < 0.05) {
    cat("✅ Reject the null: Significant difference in mean num_tools.\n")
  } else {
    cat("❌ Fail to reject the null: No significant difference in mean num_tools.\n")
  }

} else {
  cat("❌ Not enough data in one or both groups.\n")
}
```

```
## 3. F test

result_f <- var.test(
  Impact_on_Grades ~ Willing_to_Pay_for_Access,
  data = df
)

print(result_f)

if (result_f$p.value < 0.05) {
  print("Reject null: Variances are significantly different.")
} else {
  print("Fail to reject null: Variances are not significantly different.")
}



## 4. ANOVA

anova_model <- aov(Daily_Usage_Hours ~ Stream, data = df)
summary(anova_model)
p_value <- summary(anova_model)[[1]][["Pr(>F)"]][1]
if (p_value < 0.05) {
  print("Reject null: At least one Stream has different mean hours.")
} else {
  print("Fail to reject null: No significant difference among Streams.")
}


```

```

## 5. CHI SQUARE

tbl <- table(df$Stream, df$Preferred_AI_Tool)
chisq.test(tbl)
result_chi <- chisq.test(tbl)

print(result_chi)

if (result_chi$p.value < 0.05) {
  print("Reject null: Significant association between tool and willingness.")
} else {
  print("Fail to reject null: No significant association.")
}

```

TABLEAU





AI TOOL PREFERENCES & STREAM-BASED INSIGHTS
Stream (All)
Academic Streams
10
Top AI Tool
ChatGPT
Preferred AI Tools
20.9%
4.3%
23.8%
23.6%
23.1%
Bard
Claude
ChatGPT
Copilot
Gemini
Other
Professors Allow by Stream
No
Yes
Agriculture
Arts
Commerce
Engineering
Hotel-manage..
Law
Management
Medical
Pharmacy
Science
Trust in AI and Its Impact on Grades
Impact on Grades
150
100
50
0
-50
-100
AI Tool Usage Across Streams
17.1%
12.3%
12.1%
10.8%
10.1%
8.3%
7.5%
7.5%
7.4%
7.1%