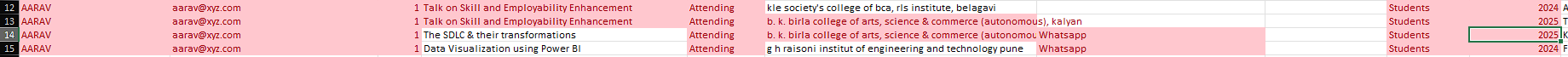
**Cleaning the Data**As I was analysing the data, I found some inconsistencies and discrepancies — the same students had registered for different events, but with varying details such as **Year of Graduation, CGPA, Experience with Python (months), City,** and even **College Name**.for e.g   
let’s look at the case of the student AARAV ,which had registered for the Event “Talk on Skill and Employability Enhancement” twice and for different events but had varying details in other fields like cgpa,city,college etc.  
  


To address these issues, I cleaned the data using Python with the Pandas library. I first removed exact duplicate entries, then grouped the records by a unique identifier (Email ID + Event) to ensure each student-event pair appeared only once. For cleaning conflicting values, I applied a mixed strategy: **for continuous fields like CGPA, I used the median** to avoid the influence of outliers, while **for categorical or discrete fields** like graduation year, experience with Python, and college name, **I used the mode (majority value)**.

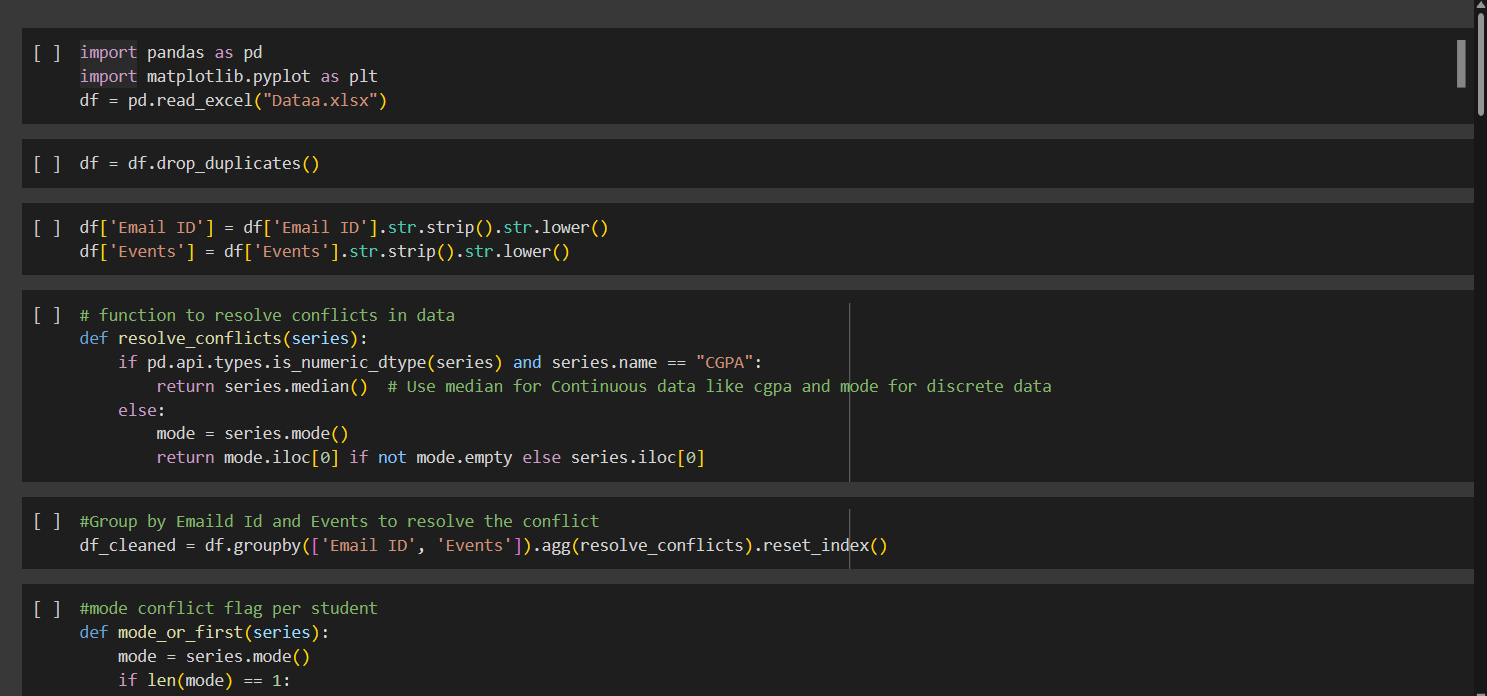
However, in cases where the mode could not be clearly determined—such as **ties between multiple values** (e.g., a student listed two different graduation years an equal number of times)—I flagged these entries with a **conflict indicator**. This helped ensure transparency in how values were resolved.

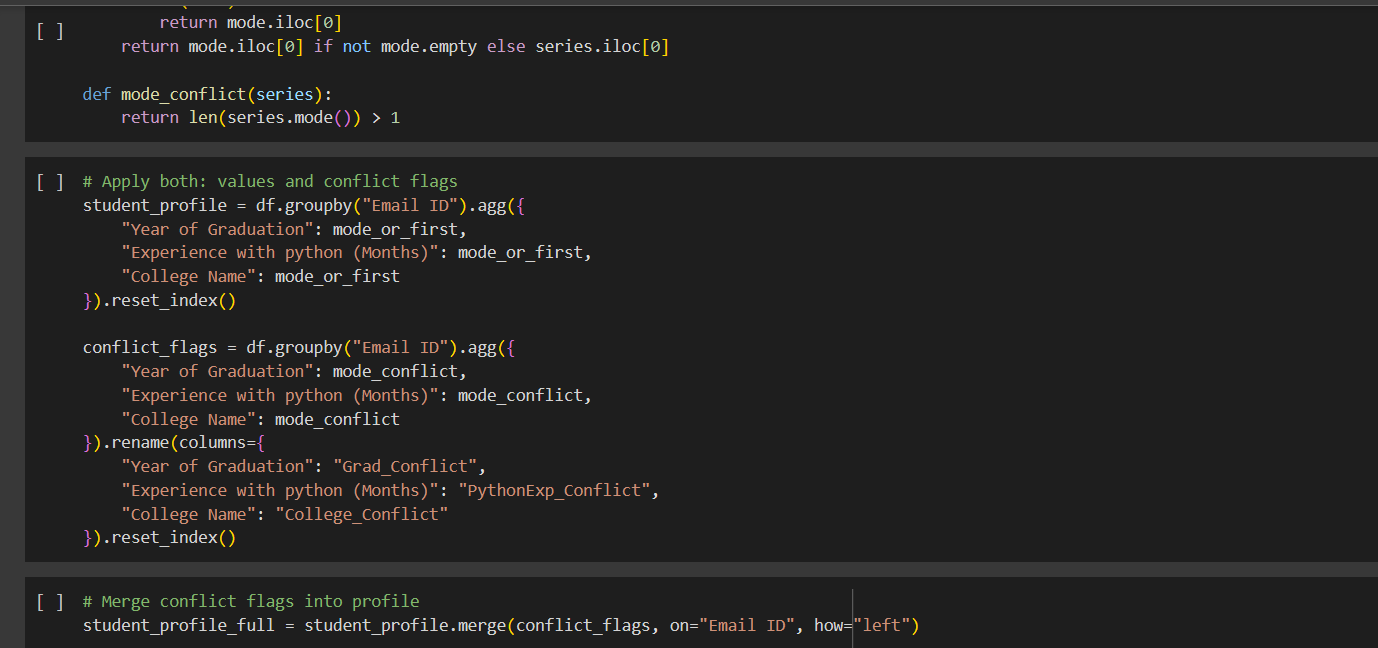
In addition to cleaning, I conducted a **data integrity check**. Students whose CGPA values varied significantly or who reported multiple cities were flagged as “untrusted” (FALSE in the Is\_Trusted column). This allowed me to retain all meaningful participation records while filtering out unreliable personal data during sensitive analyses, such as calculating average CGPA or analysing city-wise participation.

This multi-step approach ensured both data completeness and analytical integrity.So,these are the three conflict flags I created which is for Year of Graduation , Experience with Python, College name and the Is\_Trusted column which provides integrity check on data (for e.g if same student has entered different cities or cgpa range varies by 1 across entries).

A screenshot of a computer

AI-generated content may be incorrect.



 **A screen shot of a computer program

AI-generated content may be incorrect.**

**A screenshot of a computer program

AI-generated content may be incorrect.**

**Note - Participation-based analyses include all cleaned student-event records. For insights involving personal details (e.g., CGPA, city), only trusted records without major inconsistencies or conflicts were used to ensure data reliability.**

**For all student-based analysis, only entries with the designation listed as "Student" were considered. This ensures that the insights reflect the experiences and expectations of current students, excluding professionals or faculty members and others.**

Q1.How many unique students are included in the dataset?

A screenshot of a computer

AI-generated content may be incorrect.  
  
  
  
  
  
**Conclusion:** I simply converted the field of Email ID to a measure and applied the COUNTD() function to get the number of unique Email IDs, which gives us the number of unique students. I filtered the Designation field to keep only “Student” so that the count reflects student-based data only. Then I displayed the result using Text,which tell us the average count of Student is 2071

Q2. What is the average GPA of the students?   
  
A screenshot of a computer

AI-generated content may be incorrect.

**Conclusion:** I used data that is flagged as trusted and designated as Student in filter and simply converted the field of Cgpa to a measure and used aggregation function average on the field to find the average of Gpa which is ,**8.041**

Q3. What is the distribution of students across different graduation years?

A screenshot of a computer

AI-generated content may be incorrect.

**Conclusion:** Out of 2071 unique students ,582 students had conflict in Year\_of\_Graduation,thus those are excluded in this data, therefore many **Students graduating in 2023 were the most active participants** in the events.

Q4. What is the distribution of student’s experience with Python programming?  
  
A screenshot of a computer

AI-generated content may be incorrect.

**Conclusion:** Out of **2071 unique students**, **708 students had conflicting data** regarding their Python experience and were excluded.

Among the **1,363 students**, the majority had  **5 months** of Python experience, with **5 months (389)** and **3 months (284)** being the most common.  
This suggests most students are at an early learning stage or recently started programming with Python.

Q6 How does the GPA vary among different colleges?  
  
A screenshot of a computer

AI-generated content may be incorrect.

**Conclusion:** Using **trusted student data**, *St. Xavier’s College* recorded the **highest average CGPA**, followed by *New Horizon Institute*, *MIT Academy of Engineering*, *Vidyalankar Institute of Technology*, and *AP Shah Institute of Technology*. This suggests that students from these institutions performed relatively better academically among the top 5.

**Q7. Are there any outliers in the quantity (number of courses completed) attribute?**  
  
A screenshot of a computer

AI-generated content may be incorrect.

A graph with a blue line

AI-generated content may be incorrect.  
  
**Conclusion:** There are no outliers in the "Quantity" (courses completed) data.

**Q8: What is the average GPA for students from each city?**

A map of the world

AI-generated content may be incorrect.

* **South India** shows the **highest academic performance**, led by a few top-performing cities like Bengaluru. CGPAs here are strong and consistent.
* **West India** also performs very well, with several cities showing **above-average CGPAs**, close to the top range.
* **North India** has a **mix of high and average** performers. Most cities are doing well, but there’s slightly more variation.
* **East and North-East India** have **moderate but steady** performance, with CGPAs generally in the mid-8 range.
* **Central India** shows **average performance**, with limited variation and no extremes.

**Conclusion:** South and West India lead in student CGPA averages, with **Bengaluru (9.5)** and **Nashik (9.1)** leading the country in average CGPA.

**Q9: Can we identify any relationship between family income and GPA?**

**A screenshot of a computer

AI-generated content may be incorrect.**

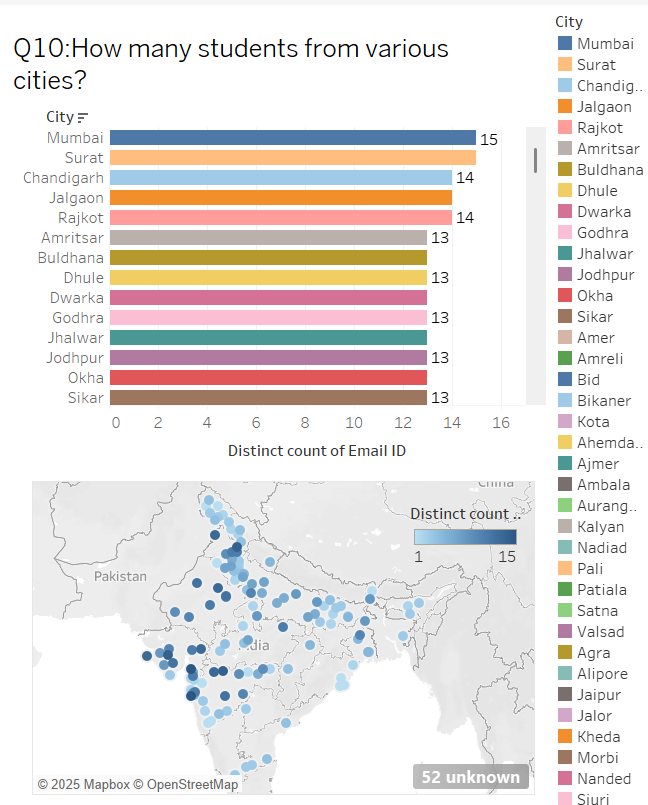
**Conclusion:** I plotted the average CGPA against ordinal family income levels to observe any trend or correlation. From the visual, it appears there isn’t a strong relationship between family income and CGPA. The average CGPA remains fairly consistent across all income groups, with only slight variation. This suggests that family income does not significantly impact academic performance, at least in this dataset. I filtered the data to include only trusted entries to ensure accuracy.

Here , Ordinal Income is

A screenshot of a computer

AI-generated content may be incorrect.

**Q10:How many students from various cities?**

****

**A screenshot of a graph

AI-generated content may be incorrect.**

**A map of the world

AI-generated content may be incorrect.**

**Conclusion:** To figure out how many students came from different cities, I filtered the data to only include records where the Designation was "Student" and kept only the trusted ones (Is Trusted = True). I then performed a COUNTD(Email ID) to make sure only unique students were counted.

This gave a clean count of distinct students per city. I plotted the data using a horizontal bar chart for easier comparison and also a map view to show geographical spread. Cities like Mumbai, Surat, and Chandigarh had the highest participation.

**Q11. How does the expected salary vary based on factors like ‘GPA’, ‘Family income’, ‘Experience with Python (Months)’?**

**A screenshot of a graph

AI-generated content may be incorrect.**

**A screenshot of a computer

AI-generated content may be incorrect.**

**A screenshot of a computer

AI-generated content may be incorrect.**

I plotted the average CGPA on the X-axis and the average expected salary (in LPA) on the Y-axis, while differentiating the data points using Family Income (color) and Experience with Python (Months) (size of the dot).

From the scatterplot and trend lines, a general positive correlation between GPA and expected salary can be observed. Students from higher-income families (like 7 LPA+) tend to have a steeper salary expectation rise with increasing GPA. Additionally, larger circles (indicating more Python experience) are often associated with higher salary expectations, suggesting experience in relevant skills also plays a role in boosting salary expectations.

In the second view, I added average values of Python Experience and Ordinal Income against CGPA in parallel axis plots, which helps visualize how these attributes trend with academic performance.

The data was filtered using “Is Trusted = True” and “PythonExp Conflict = False” to ensure clean and reliable analysis.

**Q12. Which event tends to attract more students from specific fields of study?**

**A graph of different colored bars

AI-generated content may be incorrect.**

**Conclusion:** I created a bar chart with each event on the X-axis and the distinct count of student Email IDs on the Y-axis. I applied filters to include only entries where the Designation was set to “Students” to ensure the analysis focused on the target audience.

From the visualization, the Internship Program event clearly stands out, attracting 533 students, making it the most popular event among students across fields. It is followed by Product Design & Full Stack Development (464 students) and Data Visualization using Tableau (345 students). These events, which emphasize practical industry exposure and hands-on skills, tend to draw significantly more interest than traditional or conceptual sessions.

Events such as Art of Resume Building and Talk on Skill and Employability also performed well, with over 300 students each, further highlighting students’ preference for career readiness and job-market-aligned skills.

This pattern reveals that students gravitate towards events that offer tangible career value, either through direct opportunities (like internships) or through skill enhancement closely tied to hiring processes.

**Q14. How many students are graduating by the end of 2024?**

**A screenshot of a computer

AI-generated content may be incorrect.**

I applied filters to include only entries where Designation = Students and Grad Conflict = False to ensure data consistency. I then counted the distinct student Email IDs for those with a Year of Graduation up to 2024.

From the resulting bar chart, I found that 963 students are expected to graduate by the end of 2024. This subset forms a key group for early career opportunities, internships, and job readiness programs, making them a strategic audience for skill-building and placement-related initiatives.

**Q15. Which promotion channel brings in more student participations for the event?**

**A screenshot of a graph

AI-generated content may be incorrect.**

**Conclusion:** I filtered the data to include only students with valid entries (College Conflict = False and Designation = Students) to ensure accurate representation. I then counted the number of distinct student Email IDs grouped by **Promotional Channel**.

From the chart, it's clear that **WhatsApp** emerged as the most impactful promotional channel, bringing in **341 student participants**. **Email** followed as the second most effective method with **208 participants**. Other channels like **SPOC/College Professors** (60) and **Others** (65) had a moderate reach, while platforms such as **Instagram**, **Telegram**, and **Twitter** showed minimal effectiveness, with fewer than 10 participants each.

These results reinforce the importance of **direct and instant communication platforms**, especially WhatsApp, when targeting student audiences. It also suggests that traditional outreach methods like institutional SPOCs still hold value, whereas popular social platforms may not be as effective for event promotions in this context.

**Q16.Find the total number of students who attended the events related to Data Science.**

***(From all Data Science related courses.)***

***A screenshot of a computer

AI-generated content may be incorrect.***

**Conclusion:** I filtered the data to include only students (Designation = Students) and selected events related to Data Science, including:

* Artificial Intelligence
* Is Data Science for You?
* Hello ML and DL
* Data Visualization using Tableau

I used distinct Email IDs to avoid counting duplicates and aggregated attendance across all selected events.

From the chart, a total of **929** students participated in at least one of these Data Science–related events. Individually, the most attended was **Data Visualization using Tableau with 345 participants**, followed by Hello ML and DL (259), Is Data Science for You? (215), and Artificial Intelligence (110).

This indicates strong interest among students in applied and visual aspects of data science, with hands-on tools like Tableau attracting the highest engagement.

***Top of Form***

***Bottom of Form***

**Q17. Do those who have high CGPA & more experience in language have higher expectations for salary? *(Use Average)***

**A screenshot of a computer

AI-generated content may be incorrect.**

I analyzed the relationship between CGPA (X-axis) and expected salary in LPA (Y-axis), using average values. To assess the influence of technical skills, I split the data based on Python experience (circle for high, square for low) and color-coded them into two groups:

* Blue: High CGPA & High Python Experience
* Red: Low CGPA & Low Python Experience

A screenshot of a computer

AI-generated content may be incorrect.

Only trusted data (Is Trusted = True) was considered to maintain accuracy.

From the scatterplot and trend lines, students with both high CGPA and strong Python experience clearly show higher salary expectations. The blue trend line rises steeply with increasing CGPA, indicating that as academic performance improves, so do salary aspirations—especially when coupled with relevant technical skills.

In contrast, the red group shows a much flatter trend, suggesting that low CGPA and minimal language experience lead to lower and more static salary expectations, regardless of GPA changes.

This highlights the compounded value of both academic excellence and hands-on experience in shaping career expectations.

**Q18. How many students know about the event from their colleges? Which are the Top 5 colleges?**

A screenshot of a graph

AI-generated content may be incorrect.

**Conclusion:** I plotted college-wise participation by counting the distinct student responses (based on Email IDs) for those who reported learning about the event through their college. The chart is filtered to include only students and those with no college conflict, ensuring relevance and clarity.

From the visualization, the Top 5 colleges with the highest number of students informed about the event through college channels are:

1. Vidyalankar Institute – 7 students
2. Wilson College – 6 students
3. Don Bosco College of Engineering – 5 students
4. KLE Society's College of Engineering – 5 students
5. LD College of Engineering – 4 students

These institutions seem to be more active in promoting the event internally, possibly through SPOCs, professors, or official college communications.