Name -> Pranit Gore

Domain -> Python

Project Title -> CareerPath Analytics

Start Date -> 19th June 2024

Position -> Intern

Intern ID -> IP -3928

End Date -> 4nd July 2024

INDEX

Problem Statements:

Basic Questions:

1. How many unique students are included in the dataset?
2. What is the average GPA of the students?
3. What is the distribution of students across different graduation years
4. What is the distribution of student’s experience with Python programming?
5. Are there any outliers in the ‘attending status’ and ‘quantity(number of courses completed)’ attribute?
6. How does GPA vary among different colleges?(Show top 5 results)
7. What is the average GPA for student from each city?
8. Can we identify any relationship between family income and GPA?

Moderate Questions:

1.How does expected salary vary based on factors like ‘GPA’, ‘Family income’, ‘Experience with python’?

2.Which event tend to attract more students from specific fields of study?

3.Do student in leadership positions during their college years tend to have higher GPA’s or better expected salary?

4.Is there a correlation between leadership skills and expected salary of the students?

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

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

7.Find total number of students who attended events related to Data Science?

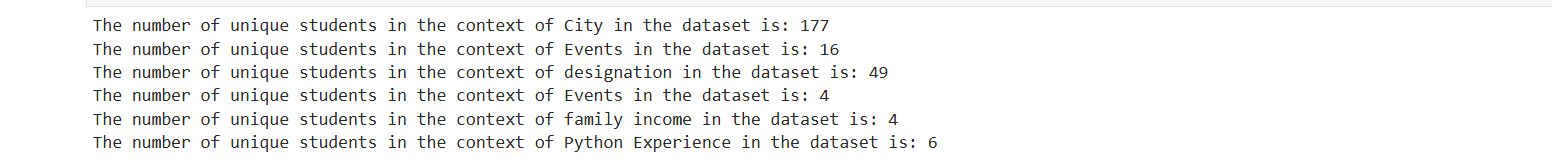
8.How many students know about the event from their college? Which of these top 5 colleges?

Qn1: How many unique students are included in the dataset?

Code:



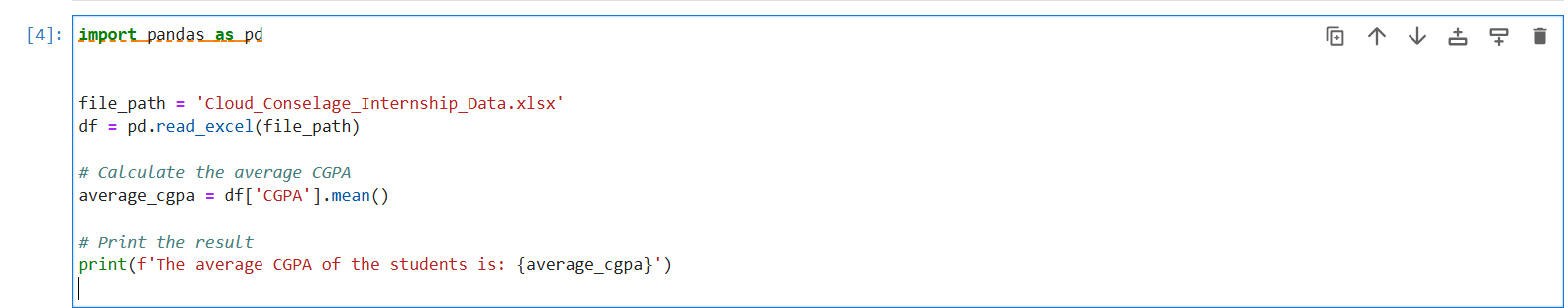
Output:

 Question 2: What is the average GPA of the students?

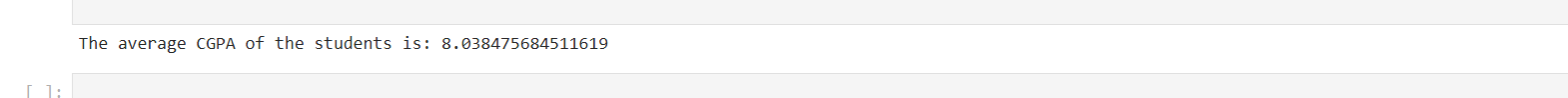
Conclusion:

We used the ‘nunique()’ function via. Pandas to get the required result. We considered separate variables for separate columns of the dataset to get better results.

Code:



Output:



Conclusion: Question 3: What is the distribution of students across different graduation years?

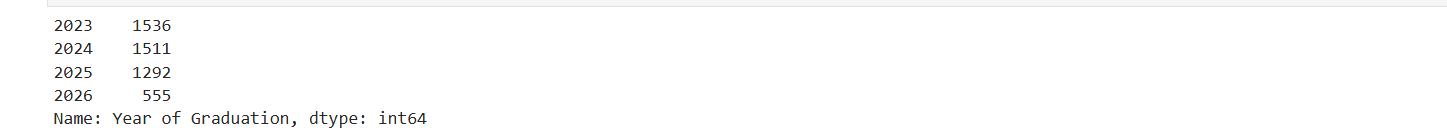
We used the ‘.mean()’ function to calculate the average GPA of the students

Code:



Output:

Output:

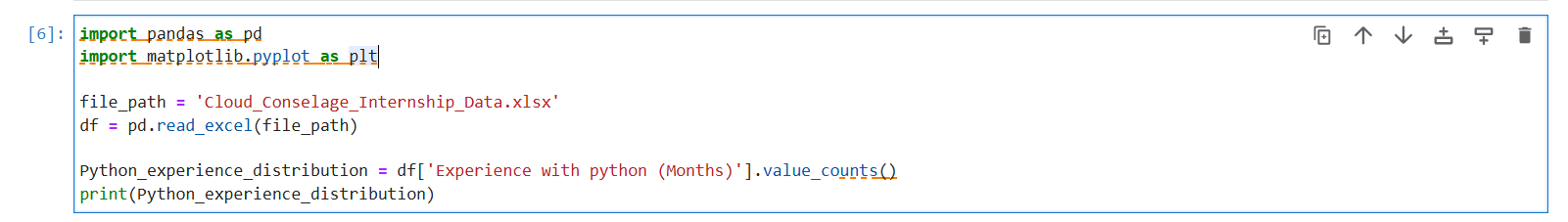


Conclusion:

We used the ‘ .value\_counts()’ function to count the number of occurrences of students in each respective graduation year.

Question 4: What is the distribution of student’s experience with Python programming?

Code :



Output:



Conclusion:

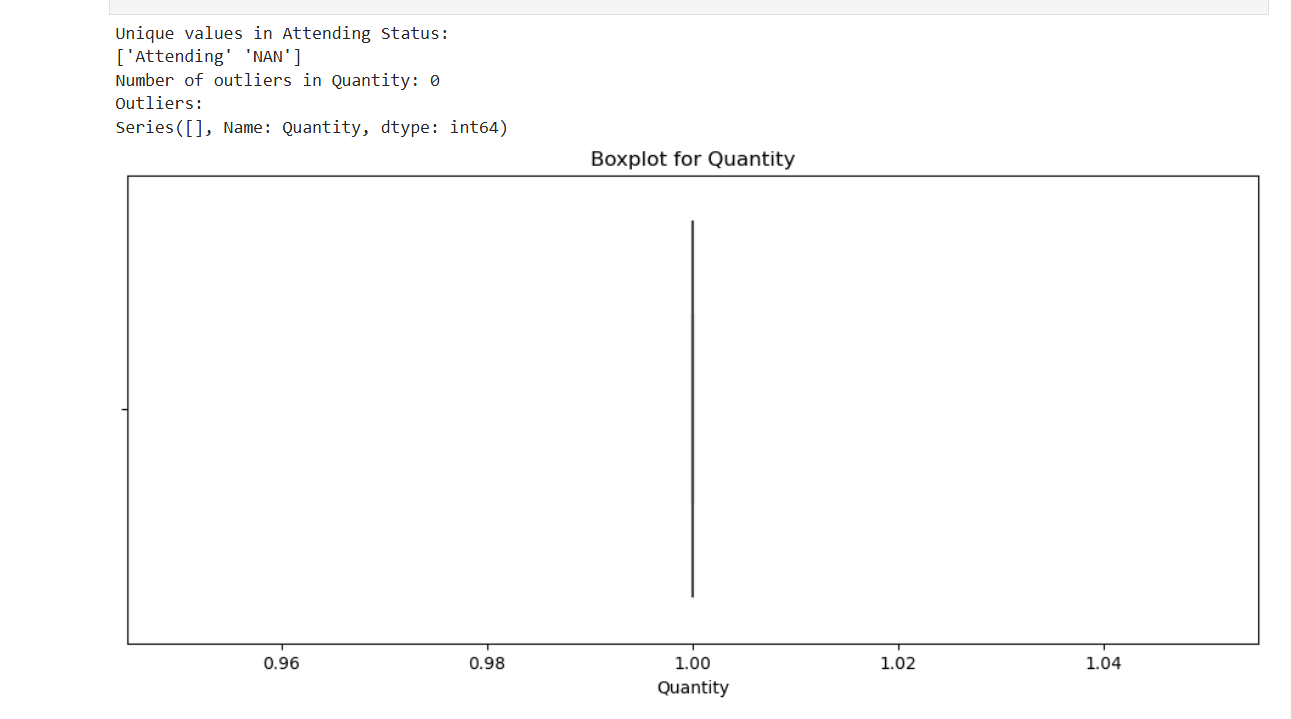
We used the ‘.value\_counts()’ function to count the number of occurrences of students of experience students have with python(in months)

Question 5: Are there any outliers in the ‘attending status’ and ‘quantity(number of courses completed)’ attribute?

Code:



Output:



Conclusion:

The code uses the Interquartile Range (IQR) method to identify outliers in the "Quantity (Number of Courses Completed)" attribute.

It effectively identifies and visualizes outliers in the dataset.

It highlights potential anomalies in the number of courses completed, which could inform further data cleaning or analysis efforts.

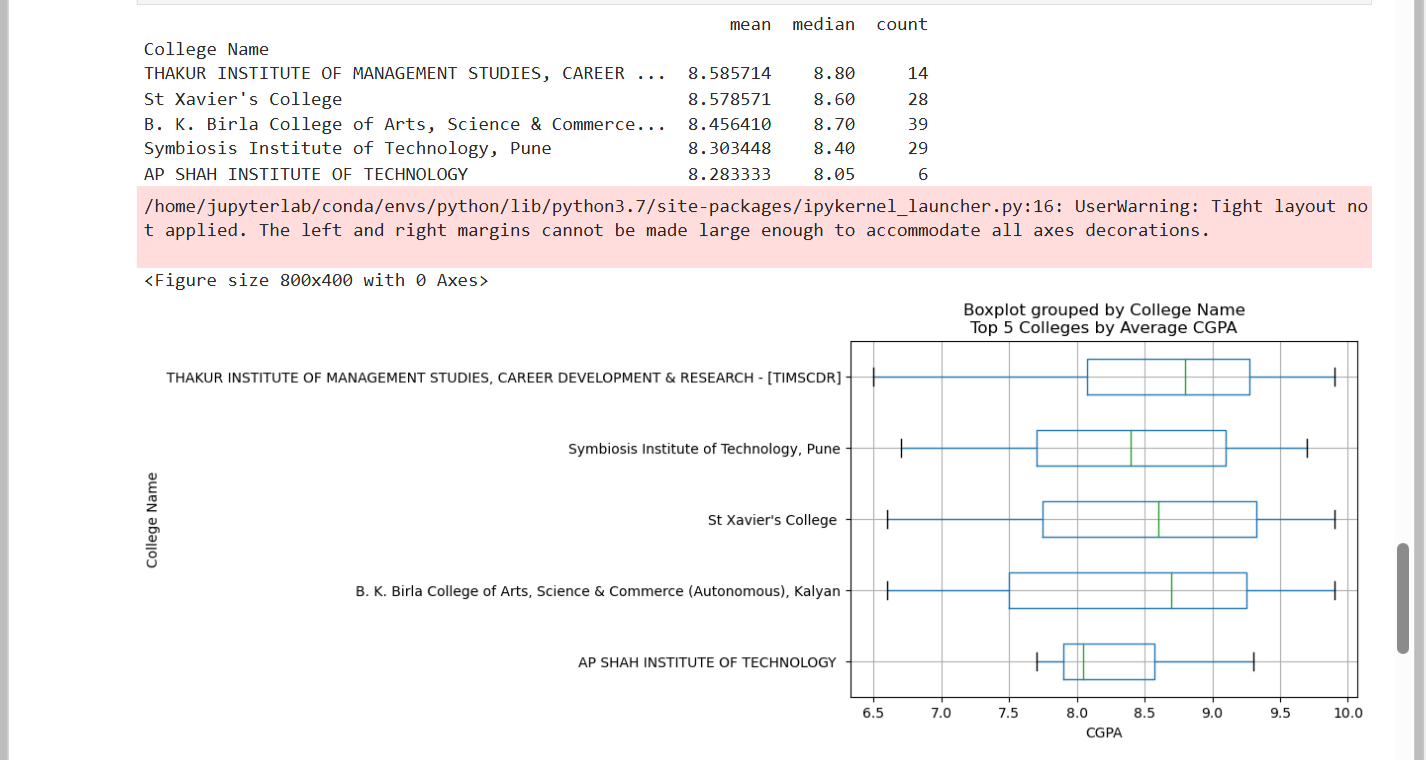
NEXT QUESTION AT NEXT PAGE

Question 6: How does GPA vary among different colleges?(Show top 5 results)

Code:



Output:



Conclusion:

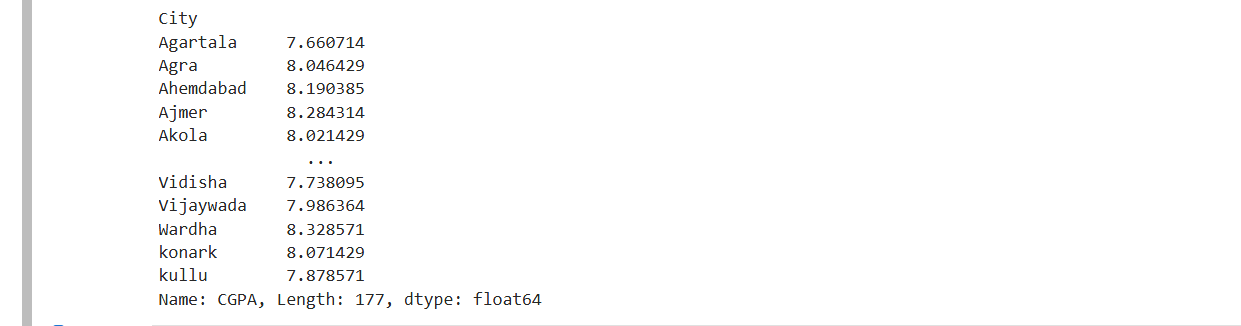
We used the ‘.groupby()’ and ‘.agg()’ function to group the college name and CGPA and find mean, median and count of them respectively. We used the ‘boxplot()’ function to create a box plot to visualize the distribution of CGPA across different colleges. With the ‘sort()’ and ‘head()’ function , we selected the top 5 colleges of the distribution and created the graphical representation of the same.

Question 7: What is the average GPA for student from each city?

Code:



Output:



Conclusion:

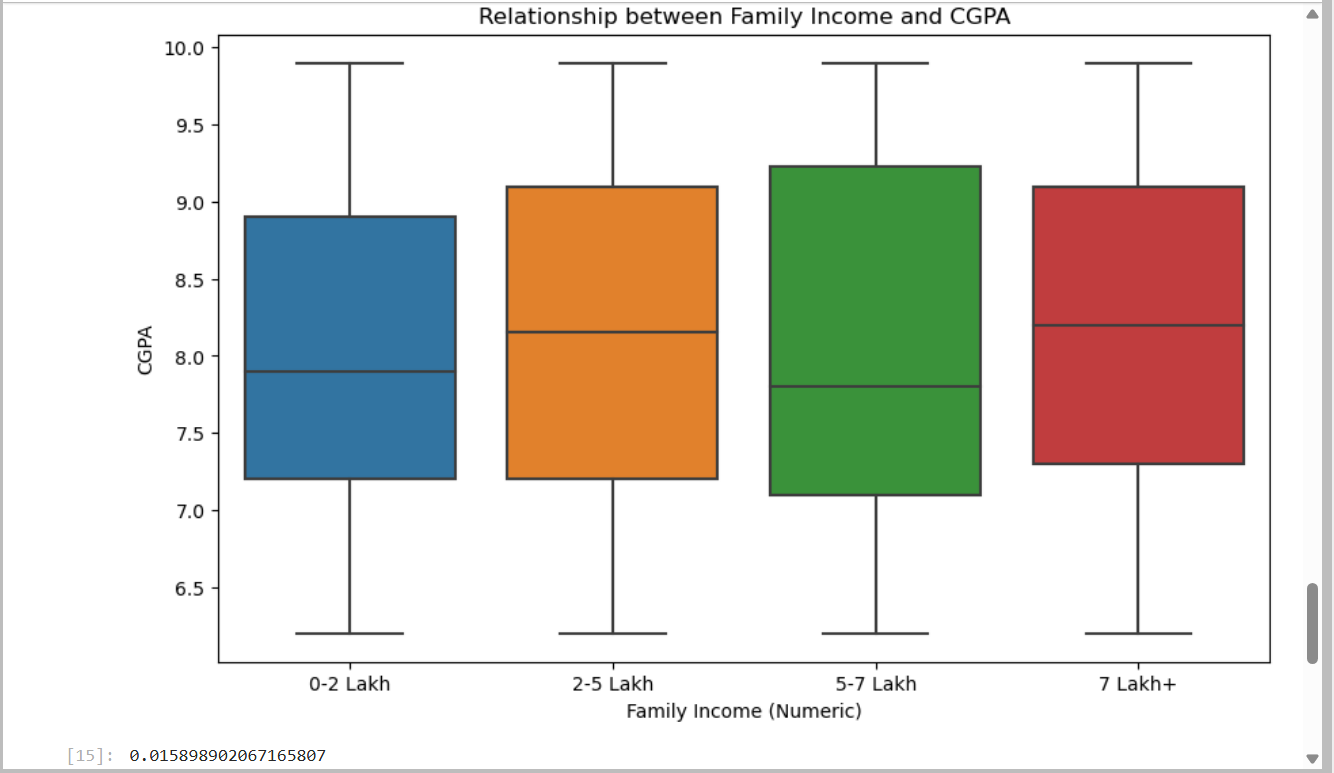
We used the ‘groupby()’ function to group the contents of city and CGPA from the datasheet and find their mean using the ‘.mean()’ function in order to calculate the average GPA of student from each city.

Question 8: Can we identify any relationship between family income and GPA?

Code:



Output:



Conclusion:

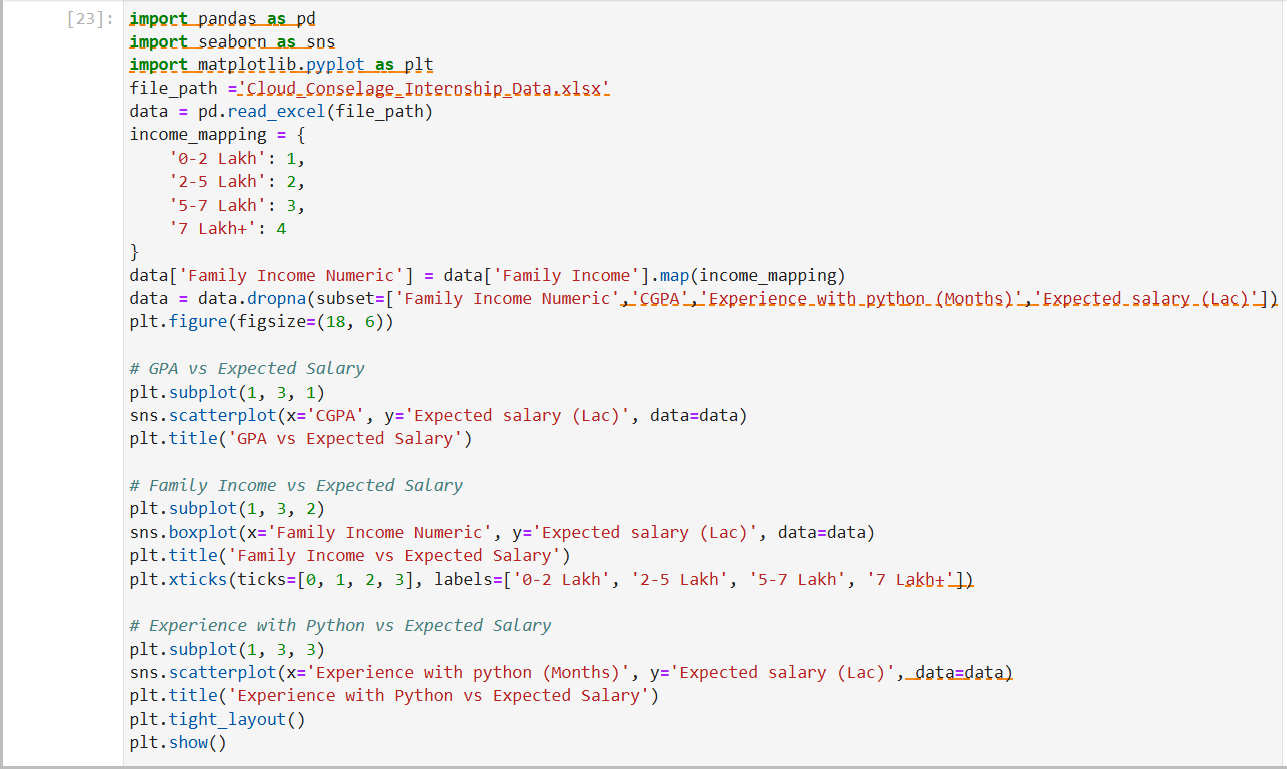
We mapped the range of family income in terms of numbers i.e. 1,2,3 and 4 because the 7 Lakh+ range is not possible to be converted to a float literal.

The correlation coefficient between family income and CGPA is approximately 0.016. This value is very close to zero, indicating a very weak positive correlation between family income and CGPA.

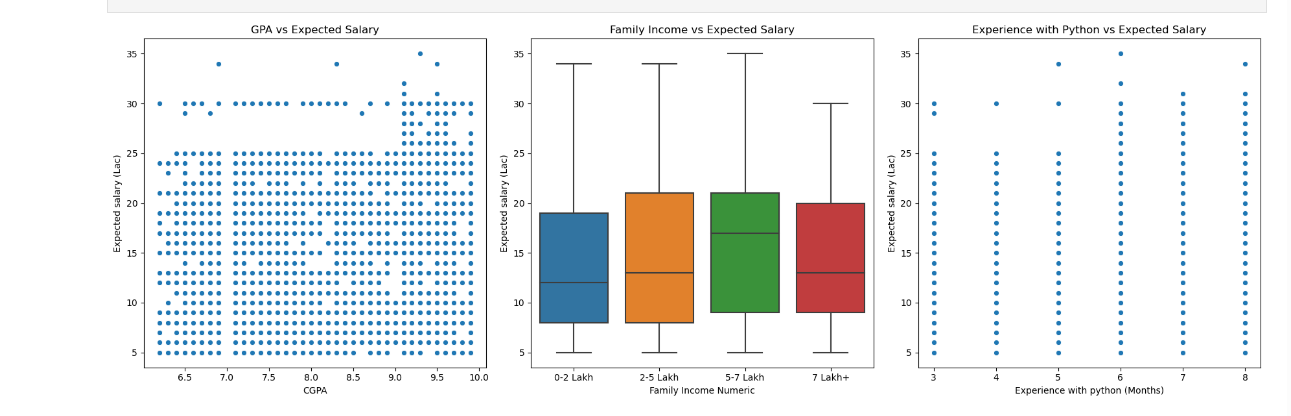
MEDIUM QUESTIONS:

Question 1: How does expected salary vary based on factors like ‘GPA’, ‘Family income’, ‘Experience with python’?

Code:



Output:



Conclusion:

**GPA vs Expected Salary**: A scatter plot showing the relationship between GPA and expected salary.

**Family Income vs Expected Salary**: A box plot showing the relationship between family income and expected salary.

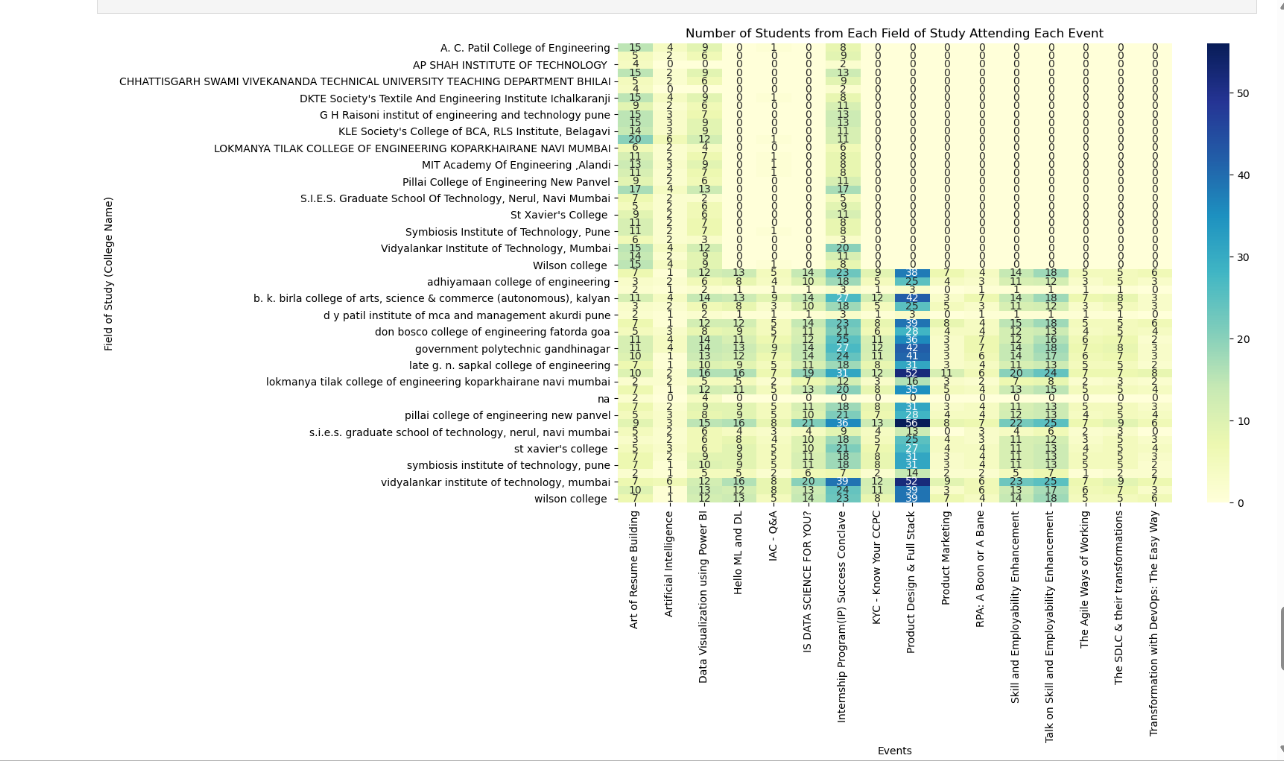
**Experience with Python vs Expected Salary**: A scatter plot showing the relationship between months of Python experience and expected salary

Question 2: Which event tend to attract more students from specific fields of study?

Code: 

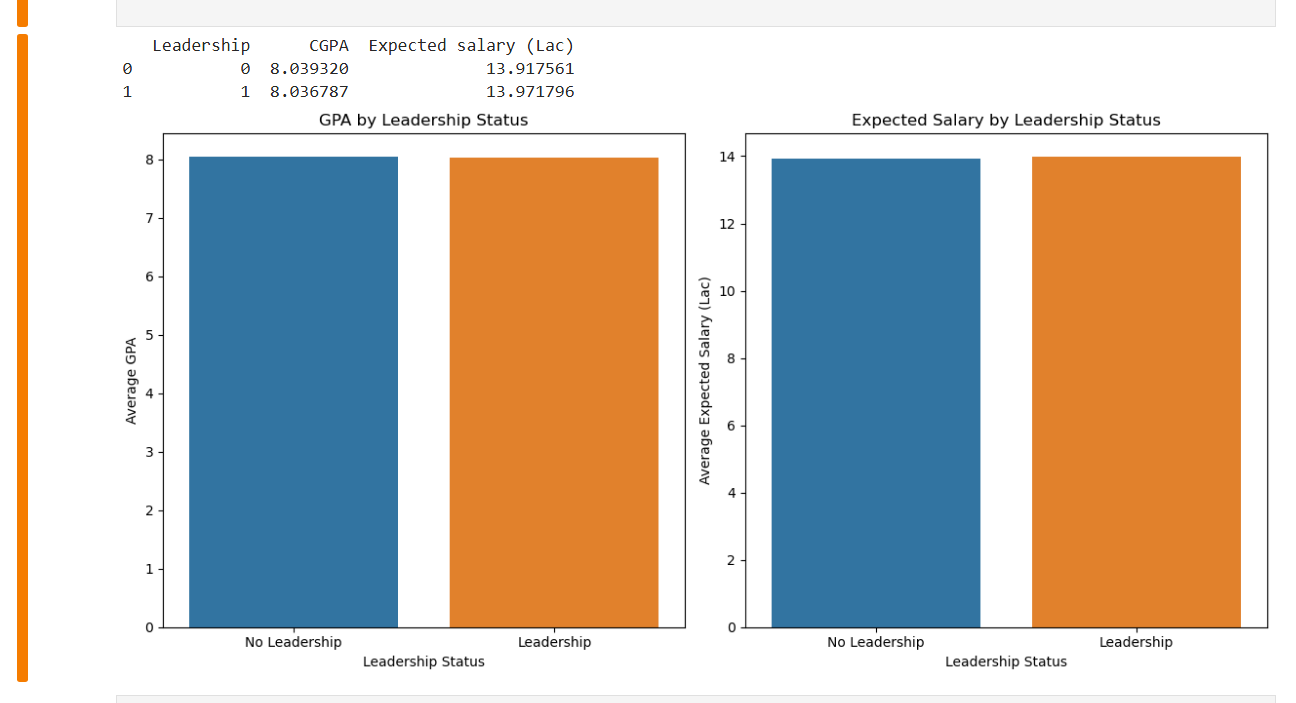
Conclusion:

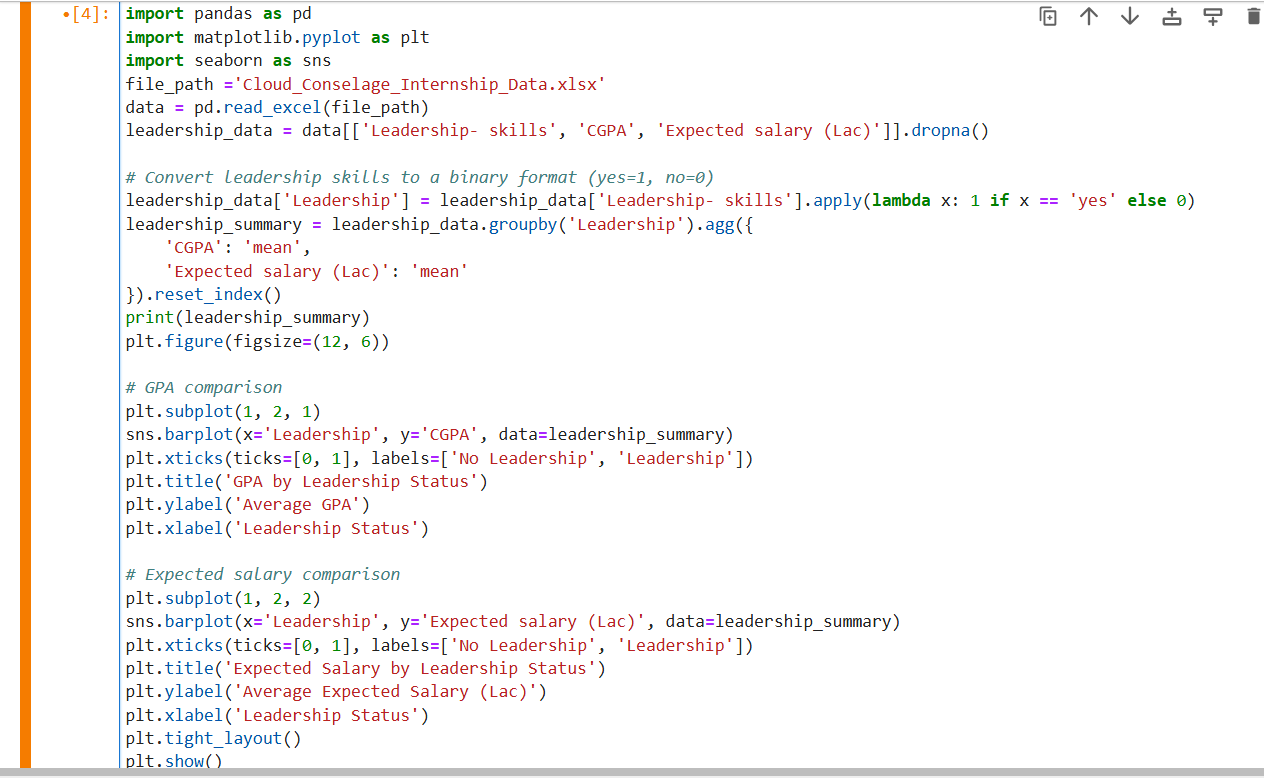
This code will generate a heatmap showing which events are popular among students from different fields of study, helping you identify trends and preferences.

Output: 

Question 3: Do students in leadership positions during their college years tend to have higher GPA’s or better expected salary?

Output:



Code: 

Conclusion:

We can conclude that if the bar plot indicates that students with leadership skills have a higher average GPA than those without, we can conclude that leadership positions tend to correlate with higher academic performance. Also, If the bar plot indicates that students with leadership skills expect a higher salary than those without, we can conclude that leadership positions tend to correlate with higher salary expectations.

Question 4: Is there a correlation between leadership skills and expected salary of the students?

Code and Output:

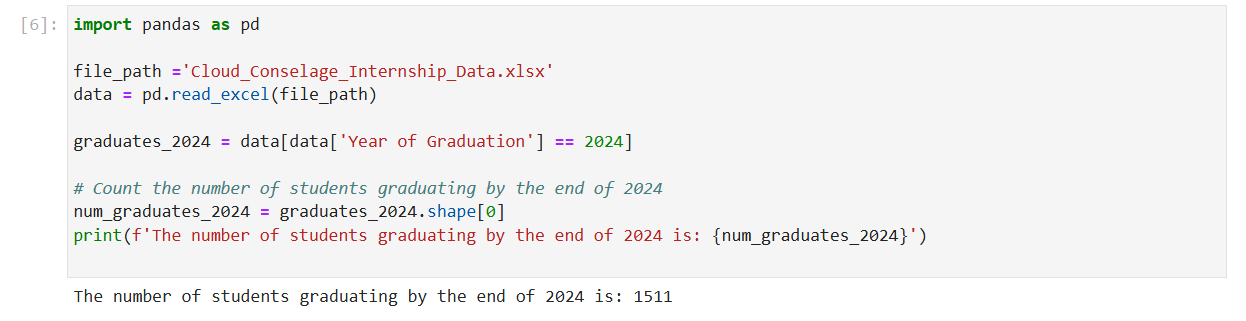


Conclusion:

The Correlation coefficient is between the values of -1 and +1. Here, +1 stands for perfect positive correlation, while -1 stands for perfect negative correlation. Note that 0 stands for no correlation.Here, the output 0.0396 indicates little to no correlation between leadership skills and expected salary.

Question 5: How many students are graduating by the end of 2024?

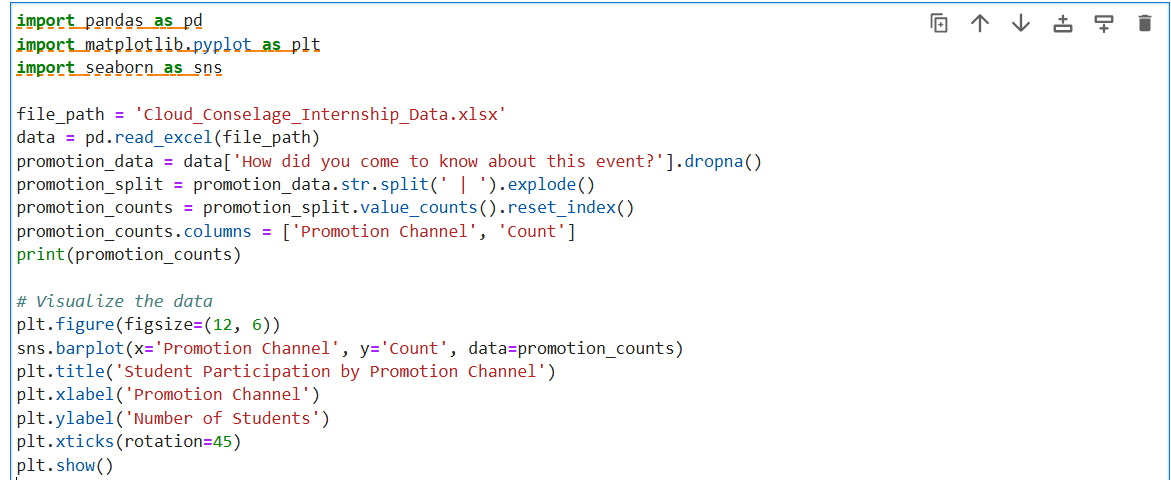
Code and Output:



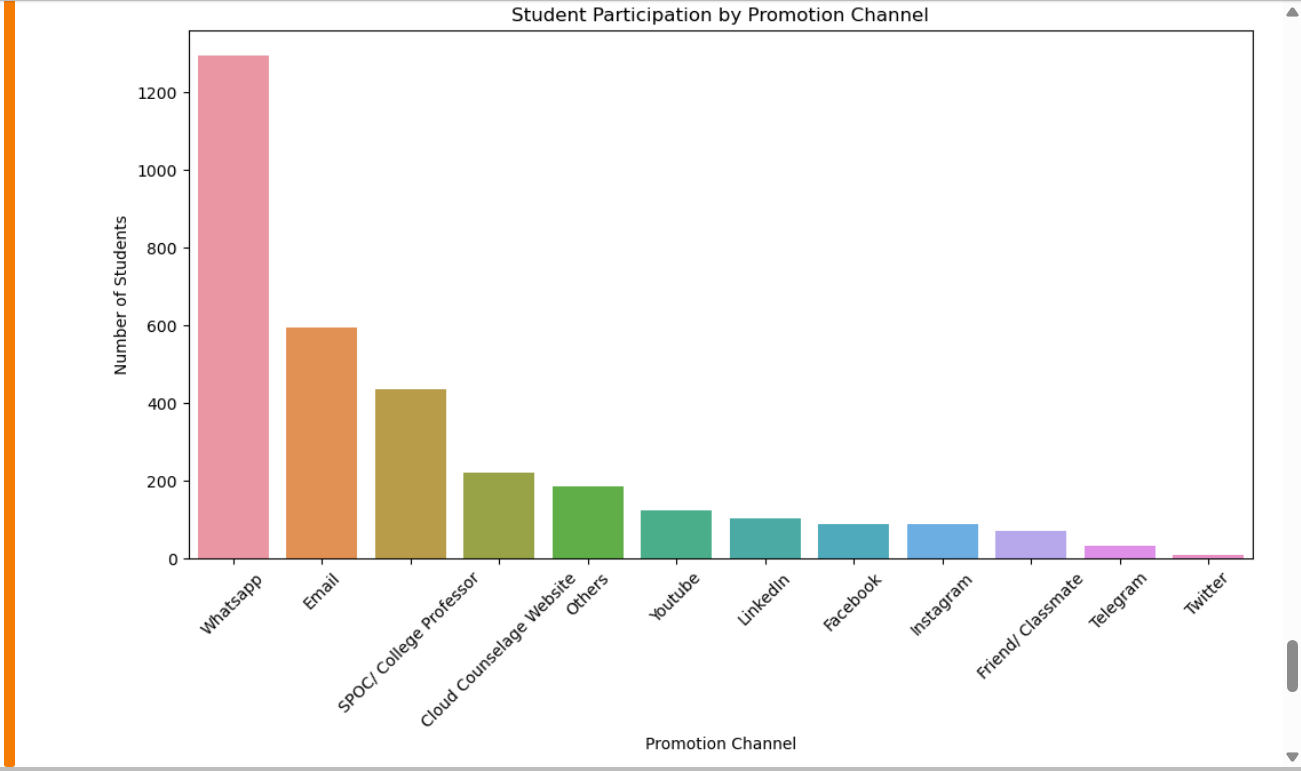
Conclusion:

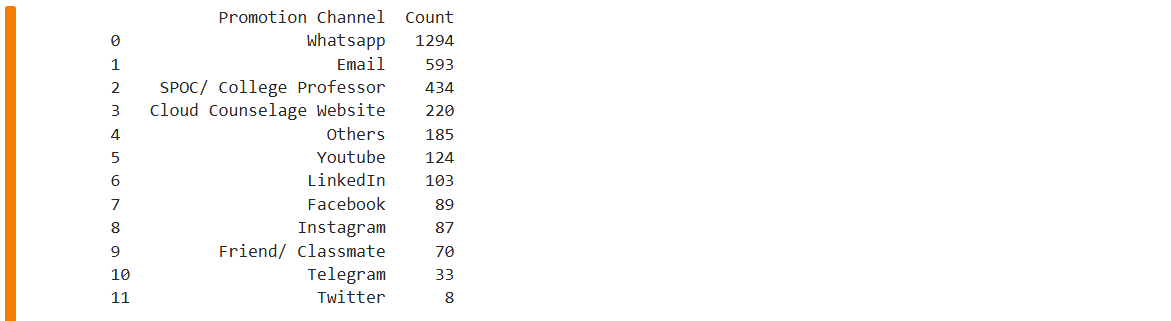
This analysis helps in understanding the graduation trend and can be useful for planning purposes, such as anticipating the number of graduates entering the job market, preparing for graduation ceremonies, or managing alumni relations.

Question 6 : Which promotion channel brings in more student participations for the event? Code:



Output:



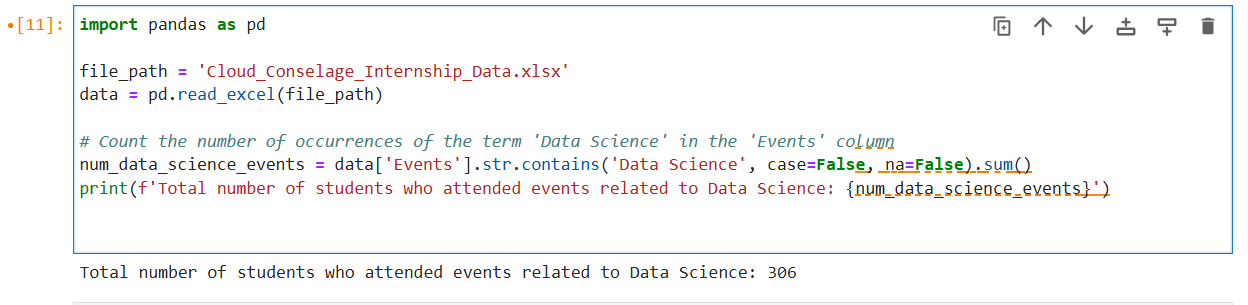


Conclusion:

The promotion channel with the highest bar represents the channel that brings in the most student participation. This helps identify the most effective promotion channel for attracting students to events. The count represents the source of promotion for the event and we can clearly see that Whatsapp was the source for promotion for majority of participants in the event.

Question 7: Find total number of students who attended events related to Data Science?

Code And Output:

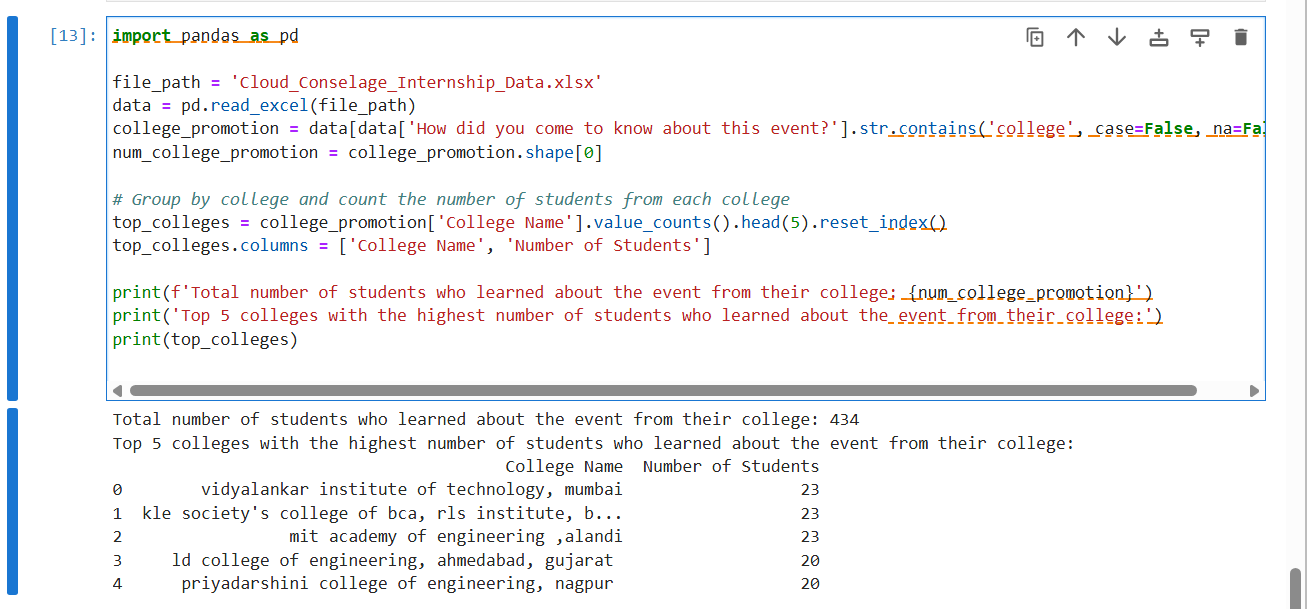


Conclusion:

We count total number of occurrences of Students enrolled in “Data Science” and output the count of occurrences using pandas.

Question 8: How many students know about the event from their college? Which of these top 5 colleges?

Code And Output:



Conclusion:

This analysis provides insight into the effectiveness of college-based promotions and identifies which colleges had the most students informed about the event through their college.

THANK YOU!