8/30/2023

Insights of Successful Students

Project Report

Submitted to Cloud Counselage Pvt. Ltd

Project Sponsor by Tushar Topale

Project Manager Harshada Topale

Ranu Bhasker
CLOUD COUNSELAGE PVT. LTD

INSIGHTS OF SUCCESSFUL STUDENTS

INTRODUCTION

In today's dynamic educational and professional landscape, understanding the factors influencing the success of student interns is crucial. This project aims to conduct a comprehensive analysis of student interns to gain insights into the relationship between their academic performance, event participation, career aspirations, and the various factors that contribute to their success. To achieve this, we have collected a comprehensive dataset containing various attributes for each student intern.

PROJECT OBJECTIVES

1. Analyze the academic performance of student interns, including GPA, course performance, and relevant metrics.
2. Investigate the extent of event participation by student interns in extracurricular and professional development activities.
3. Examine the career aspirations and goals of student interns, including their desired career paths and industries.
4. Identify and assess the factors that influence the success of student interns, such as mentorship, support systems, and personal attributes.

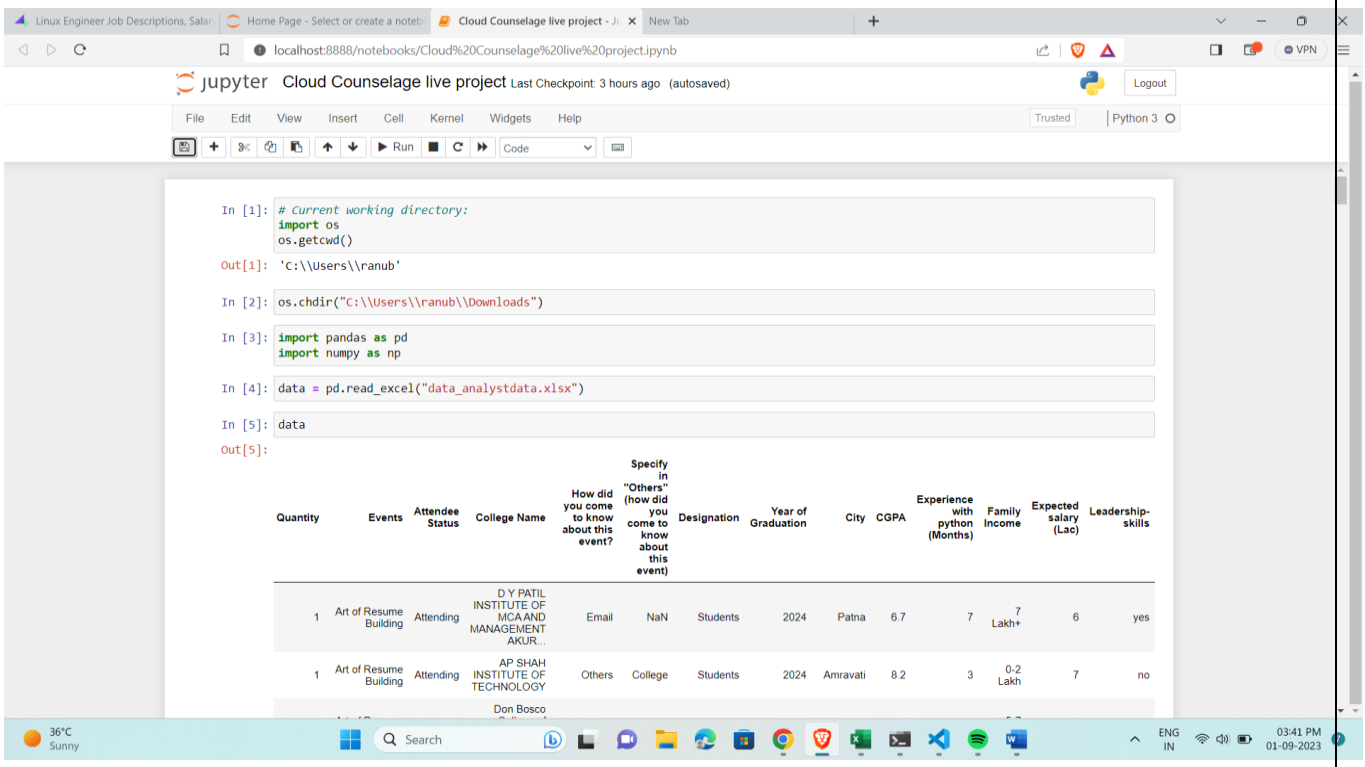
DATA COLLECTION

To achieve our objectives, we collected a dataset containing the following attributes for each student intern:

1. **Academic Performance Metrics**: This includes GPA, course grades, and academic achievements.
2. **Event Participation**: Information about the student's involvement in events, clubs, workshops, and seminars.
3. **Career Aspirations**: Data on the student's desired career path, industry preferences, and long-term goals.
4. **Factors Influencing Success**: Details about mentorship, support systems, and personal attributes like resilience, adaptability, and interpersonal skills.

METHODOLOGY

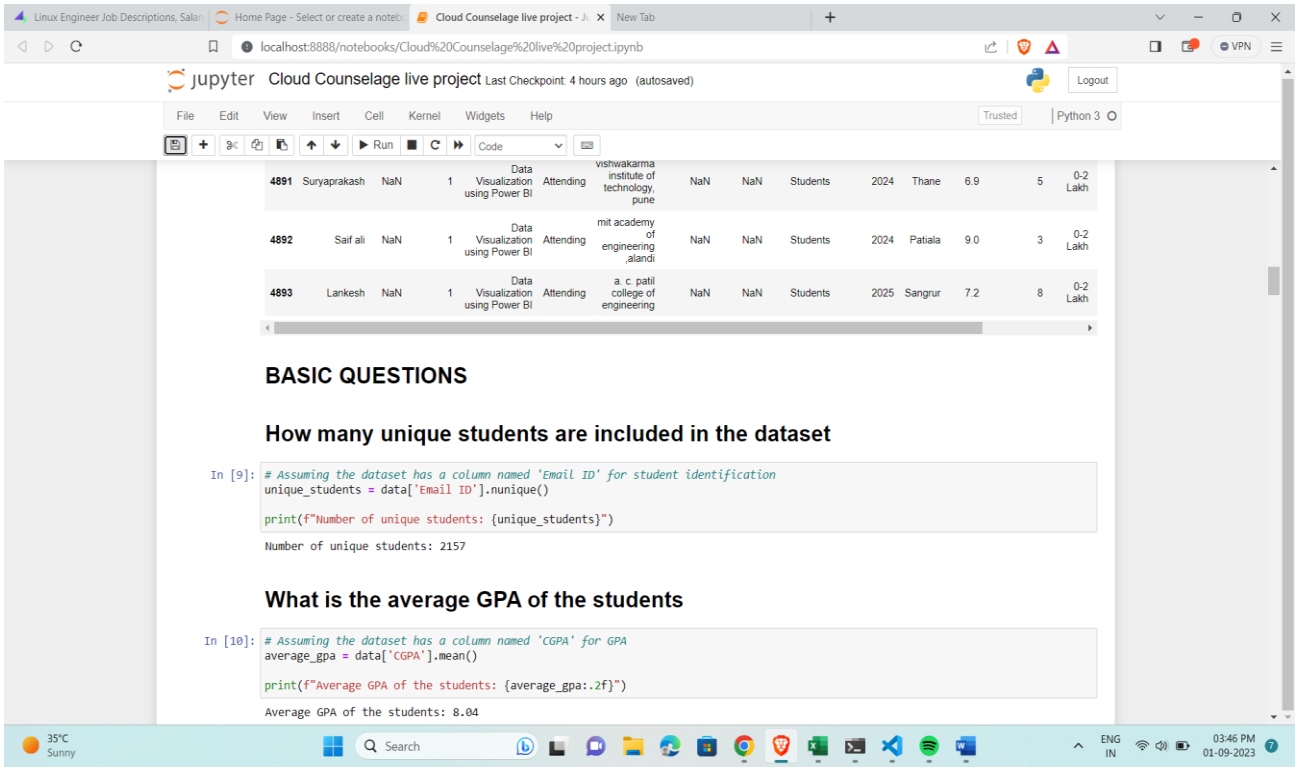
1. We conducted a thorough analysis of the dataset using statistical and data analysis techniques, including data visualization, correlation analysis, and regression modelling. This allowed us to identify trends, patterns, and relationships between different variables.
2. Firstly, we need to cleaning the data on Excel. Remove the duplicates values, and also clean or remove the unnecessary data from our datasets. Then transform the data.
3. After we need to extract the meaningful insights from the data set. We need to visualize the data with the help of python or another visualization tools.
4. Plot the trends or pattern on BI tool, create the attractive dashboard or report for better understanding of stakeholders and other team members.
5. I followed the python libraires for data visualizations.



RESULTS AND FINDINGS

Our analysis revealed several significant findings:

1. A positive correlation between academic performance and participation in extracurricular activities, suggesting that active involvement in events positively influences GPA.
2. Student interns with clearly defined career aspirations are more likely to actively engage in career-related events and workshops.
3. Mentorship and support from faculty and professionals play a vital role in student intern success.
4. Personal attributes like resilience and adaptability are strong predictors of success among student interns.
5. I have got the awesome results. I have attached the screenshots of results and findings for better understanding.



4891	Suryaprakash	NaN	1	Data Visualization using Power BI	Attending	vishwakarma institute of technology, pune	NaN	NaN	Students	2024	Thane	6.9	5	0-2 Lakh
4892	Saif ali	NaN	1	Data Visualization using Power BI	Attending	mit academy of engineering ,alandi	NaN	NaN	Students	2024	Patiala	9.0	3	0-2 Lakh
4893	Lankesh	NaN	1	Data Visualization using Power BI	Attending	a. c. patil college of engineering	NaN	NaN	Students	2025	Sangrur	7.2	8	0-2 Lakh

BASIC QUESTIONS

How many unique students are included in the dataset

```
In [9]: # Assuming the dataset has a column named 'Email ID' for student identification
unique_students = data['Email ID'].nunique()

print(f"Number of unique students: {unique_students}")
```

```
Number of unique students: 2157
```

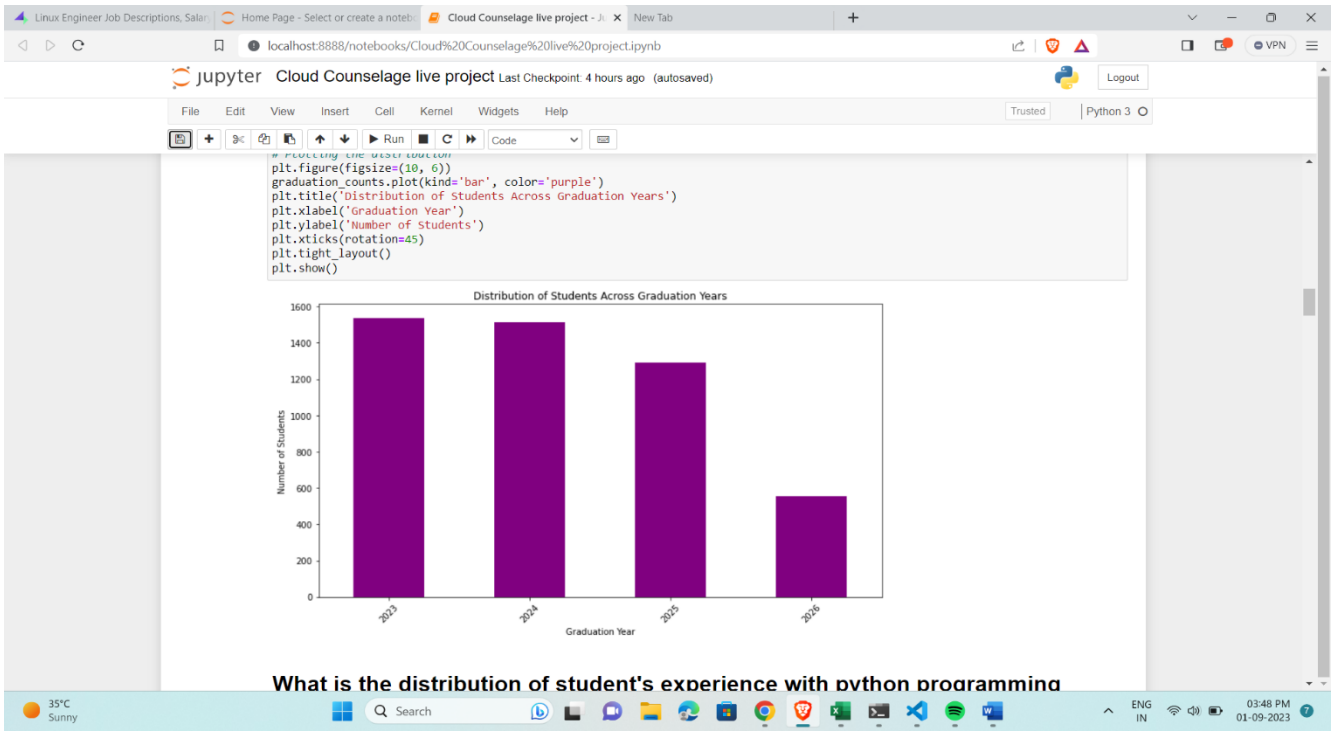
What is the average GPA of the students

```
In [10]: # Assuming the dataset has a column named 'CGPA' for GPA
average_gpa = data['CGPA'].mean()

print(f"Average GPA of the students: {average_gpa:.2f}")
```

```
Average GPA of the students: 8.04
```

It is first two questions Visualization's.



```
plt.figure(figsize=(10, 6))
graduation_counts.plot(kind='bar', color='purple')
plt.title('Distribution of Students Across Graduation Years')
plt.xlabel('Graduation Year')
plt.ylabel('Number of Students')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```


What is the distribution of student's experience with python programming

Linux Engineer Job Descriptions, Salar | Home Page - Select or create a noteb | Cloud Counselage live project - J | New Tab

localhost:8888/notebooks/Cloud%20Counselage%20live%20project.ipynb

jupyter Cloud Counselage live project Last Checkpoint: 4 hours ago (autosaved)

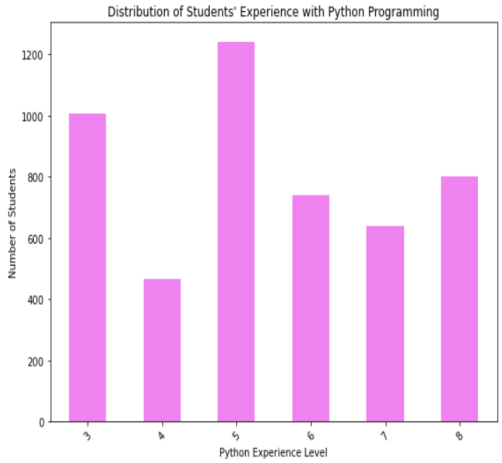
Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3

Run Code

```
# Plotting the distribution
plt.figure(figsize=(8, 6))
experience_counts.plot(kind='bar', color='violet')
plt.title("Distribution of Students' Experience with Python Programming")
plt.xlabel('Python Experience Level')
plt.ylabel('Number of Students')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

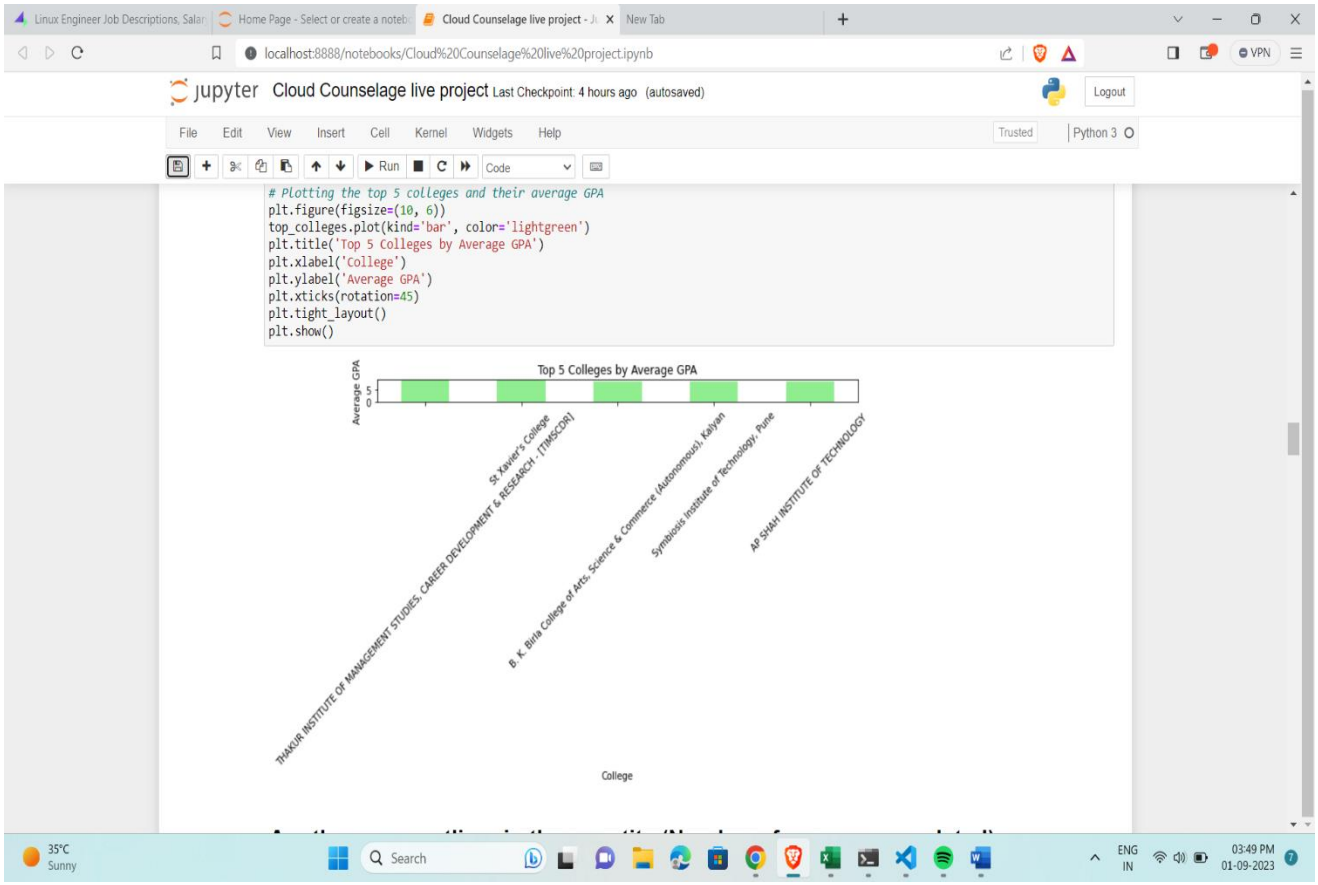


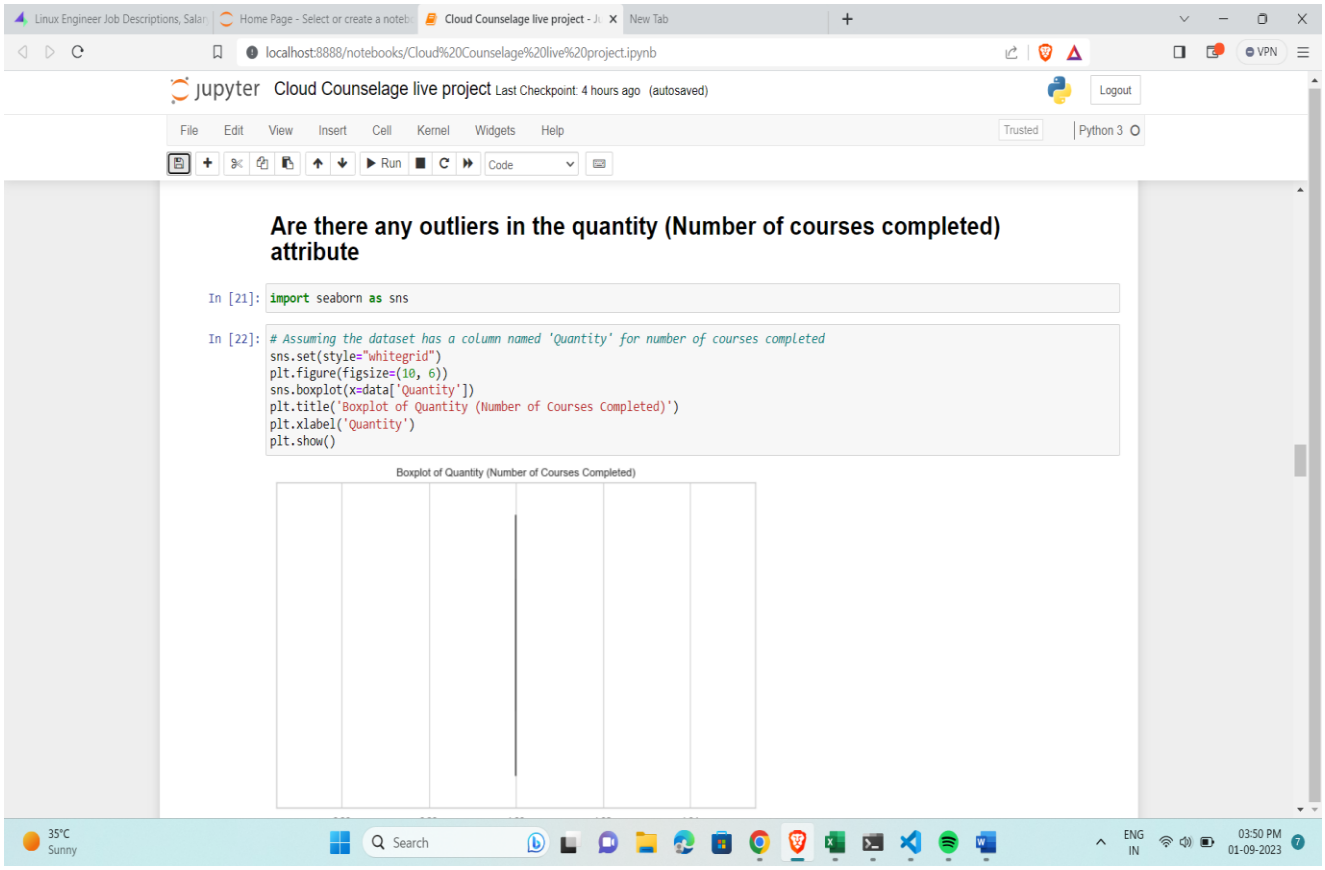
35°C Sunny

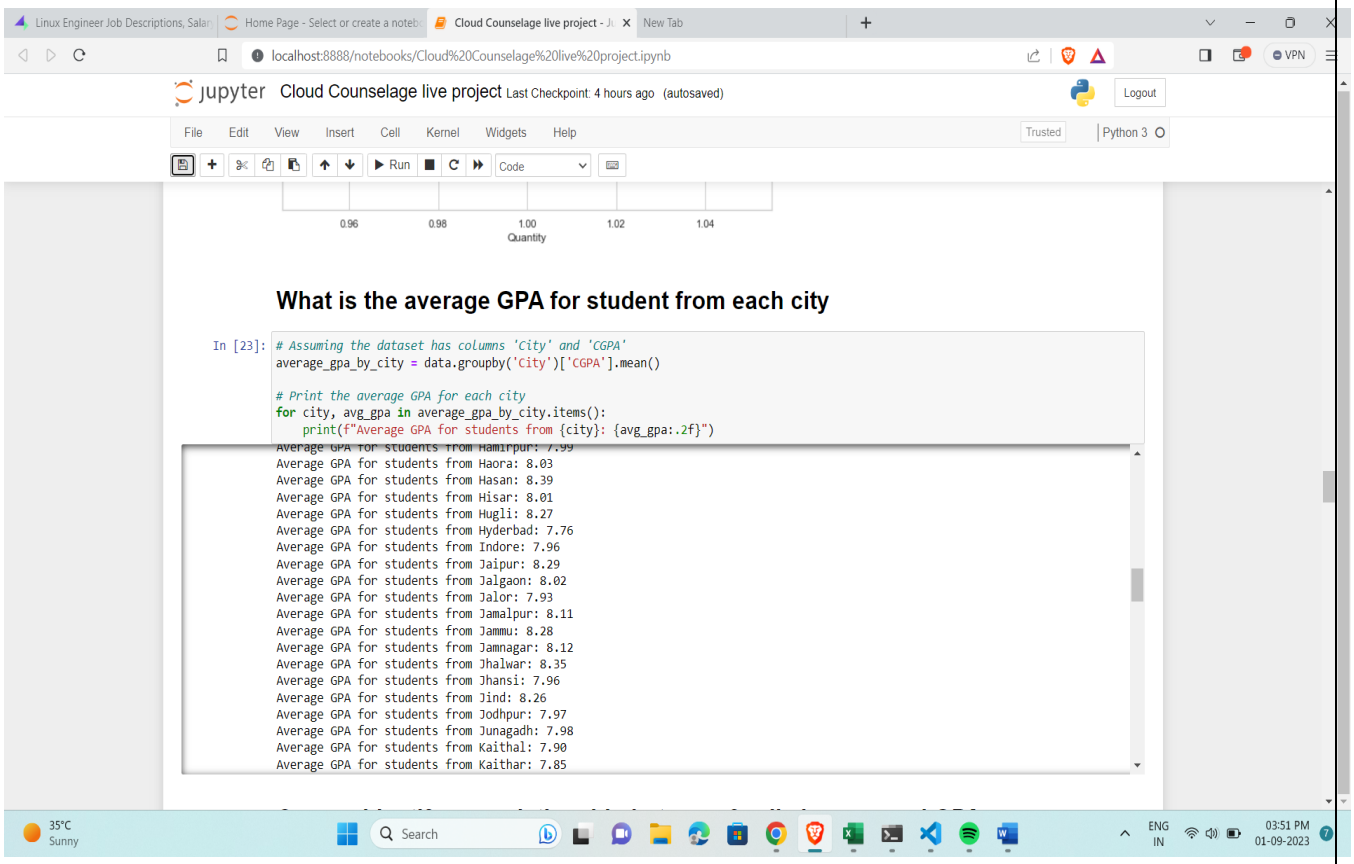
Search

ENG IN

03:48 PM 01-09-2023


Linux Engineer Job Descriptions, Salar
Home Page - Select or create a noteb
Cloud Counselage live project - J
New Tab
localhost:8888/notebooks/Cloud%20Counselage%20live%20project.ipynb
jupyter Cloud Counselage live project Last Checkpoint: 4 hours ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
# Plotting the top 5 colleges and their average GPA
plt.figure(figsize=(10, 6))
top_colleges.plot(kind='bar', color='lightgreen')
plt.title('Top 5 Colleges by Average GPA')
plt.xlabel('College')
plt.ylabel('Average GPA')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
Top 5 Colleges by Average GPA
Average GPA
5
0
THAKUR INSTITUTE OF MANAGEMENT STUDIES, CAREER DEVELOPMENT & RESEARCH - [TIMSCDR]
St. Xavier's College
B. K. Birla College of Arts, Science & Commerce (Autonomous), Kalyan
Symbiosis Institute of Technology, Pune
AP SHAH INSTITUTE OF TECHNOLOGY
College
35°C
Sunny
Search
ENG
IN
03:49 PM
01-09-2023


Linux Engineer Job Descriptions, Salar
Home Page - Select or create a noteb
Cloud Counselage live project - J
New Tab
localhost:8888/notebooks/Cloud%20Counselage%20live%20project.ipynb
jupyter Cloud Counselage live project Last Checkpoint: 4 hours ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
Are there any outliers in the quantity (Number of courses completed) attribute
In [21]:
import seaborn as sns
In [22]:
# Assuming the dataset has a column named 'Quantity' for number of courses completed
sns.set(style="whitegrid")
plt.figure(figsize=(10, 6))
sns.boxplot(x=data['Quantity'])
plt.title('Boxplot of Quantity (Number of Courses Completed)')
plt.xlabel('Quantity')
plt.show()
Boxplot of Quantity (Number of Courses Completed)
35°C
Sunny
Search
ENG
IN
03:50 PM
01-09-2023


Linux Engineer Job Descriptions, Salar
Home Page - Select or create a noteb
Cloud Counselage live project - J
New Tab
localhost:8888/notebooks/Cloud%20Counselage%20live%20project.ipynb
jupyter Cloud Counselage live project Last Checkpoint: 4 hours ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
0.96
0.98
1.00
1.02
1.04
Quantity
What is the average GPA for student from each city
In [23]:
# Assuming the dataset has columns 'City' and 'CGPA'
average_gpa_by_city = data.groupby('City')['CGPA'].mean()
# Print the average GPA for each city
for city, avg_gpa in average_gpa_by_city.items():
print(f"Average GPA for students from {city}: {avg_gpa:.2f}")
Average GPA for students from Haora: 8.03
Average GPA for students from Hasan: 8.39
Average GPA for students from Hisar: 8.01
Average GPA for students from Hugli: 8.27
Average GPA for students from Hyderbad: 7.76
Average GPA for students from Indore: 7.96
Average GPA for students from Jaipur: 8.29
Average GPA for students from Jalgaon: 8.02
Average GPA for students from Jalor: 7.93
Average GPA for students from Jamalpur: 8.11
Average GPA for students from Jammu: 8.28
Average GPA for students from Jamnagar: 8.12
Average GPA for students from Jhalwar: 8.35
Average GPA for students from Jhansi: 7.96
Average GPA for students from Jind: 8.26
Average GPA for students from Jodhpur: 7.97
Average GPA for students from Junagadh: 7.98
Average GPA for students from Kaithal: 7.90
Average GPA for students from Kaithar: 7.85
35°C
Sunny
Search
ENG
IN
03:51 PM
01-09-2023

Linux Engineer Job Descriptions, Salar | Home Page - Select or create a noteb | Cloud Counselage live project - Ju | New Tab

localhost:8888/notebooks/Cloud%20Counselage%20live%20project.ipynb

jupyter Cloud Counselage live project Last Checkpoint: 4 hours ago (autosaved)

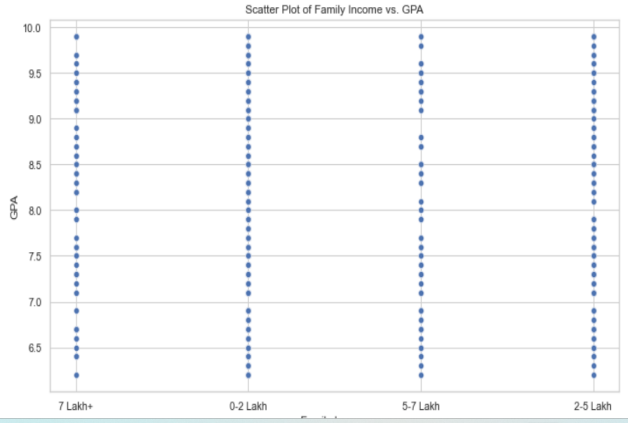
Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

Can we identify any relationship between family income and GPA

In [24]:
```
# Assuming the dataset has columns 'Family Income' and 'CGPA'
plt.figure(figsize=(10, 6))
sns.scatterplot(x='Family Income', y='CGPA', data=data)
plt.title('Scatter Plot of Family Income vs. GPA')
plt.xlabel('Family Income')
plt.ylabel('GPA')
plt.tight_layout()
plt.show()
```



35°C Sunny | Search | ENG IN | 03:52 PM 01-09-2023

Linux Engineer Job Descriptions, Salar | Home Page - Select or create a noteb | Cloud Counselage live project - Ju | New Tab

localhost:8888/notebooks/Cloud%20Counselage%20live%20project.ipynb

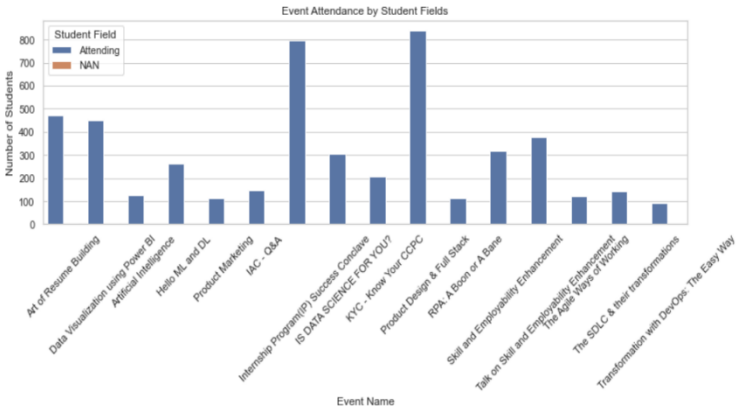
jupyter Cloud Counselage live project Last Checkpoint: 4 hours ago (autosaved)

Logout

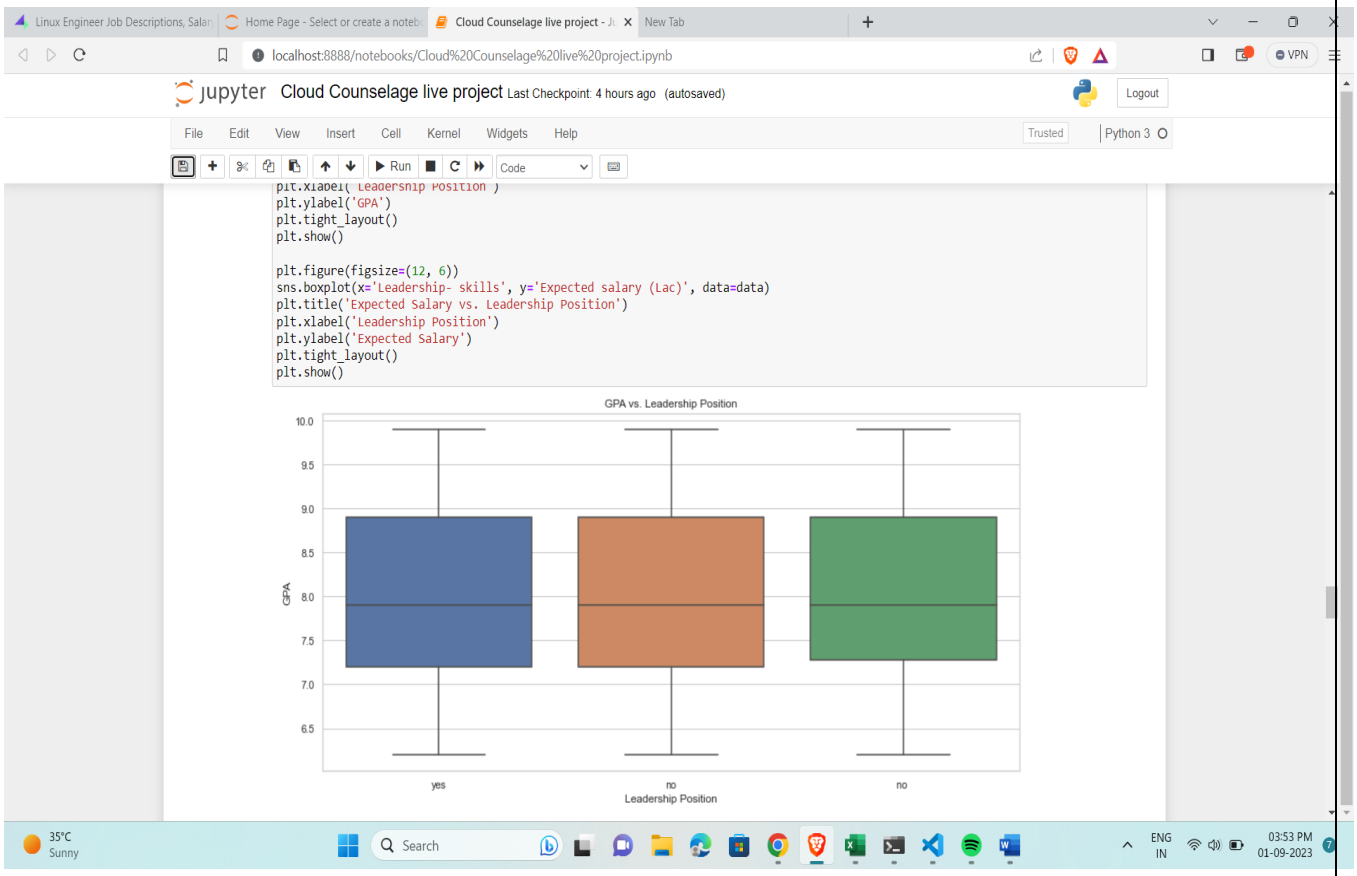
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

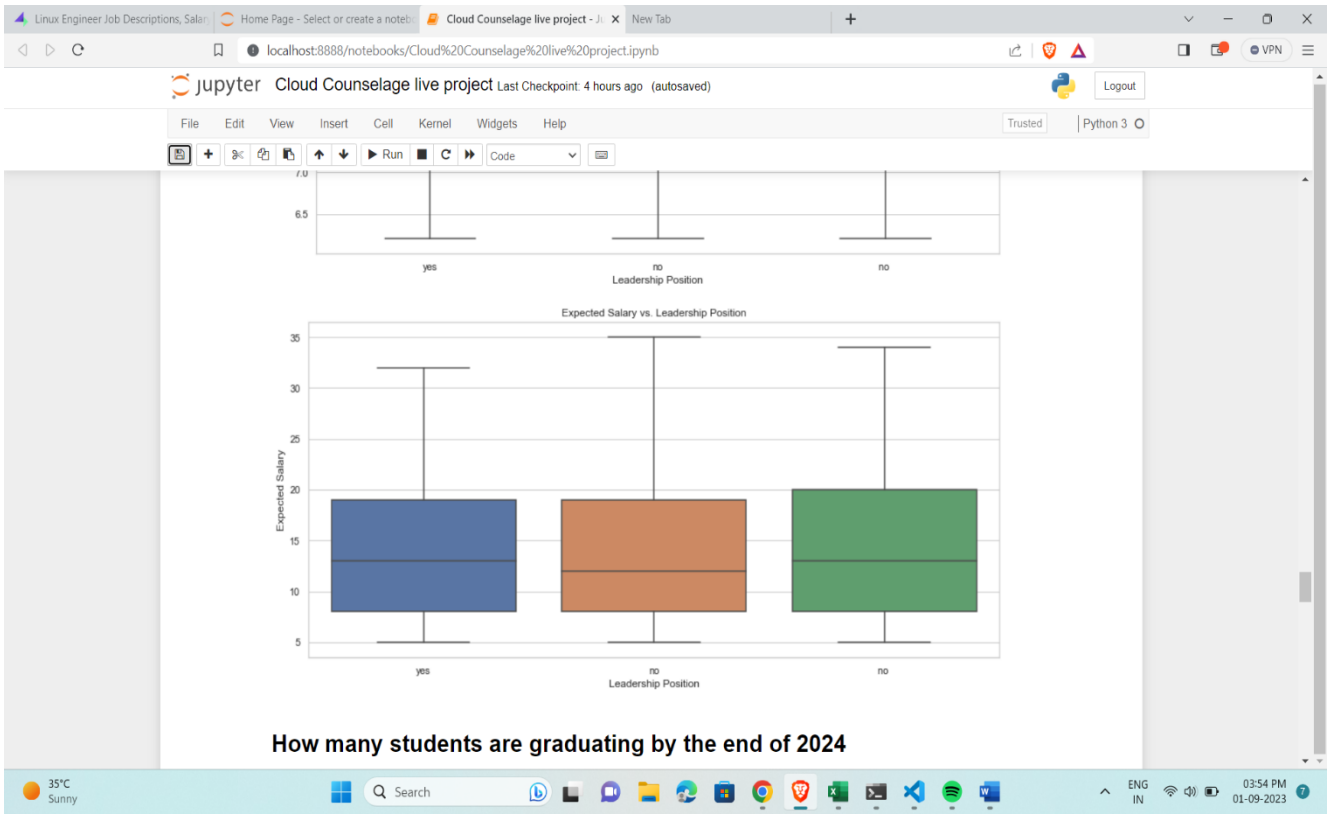
Run Code

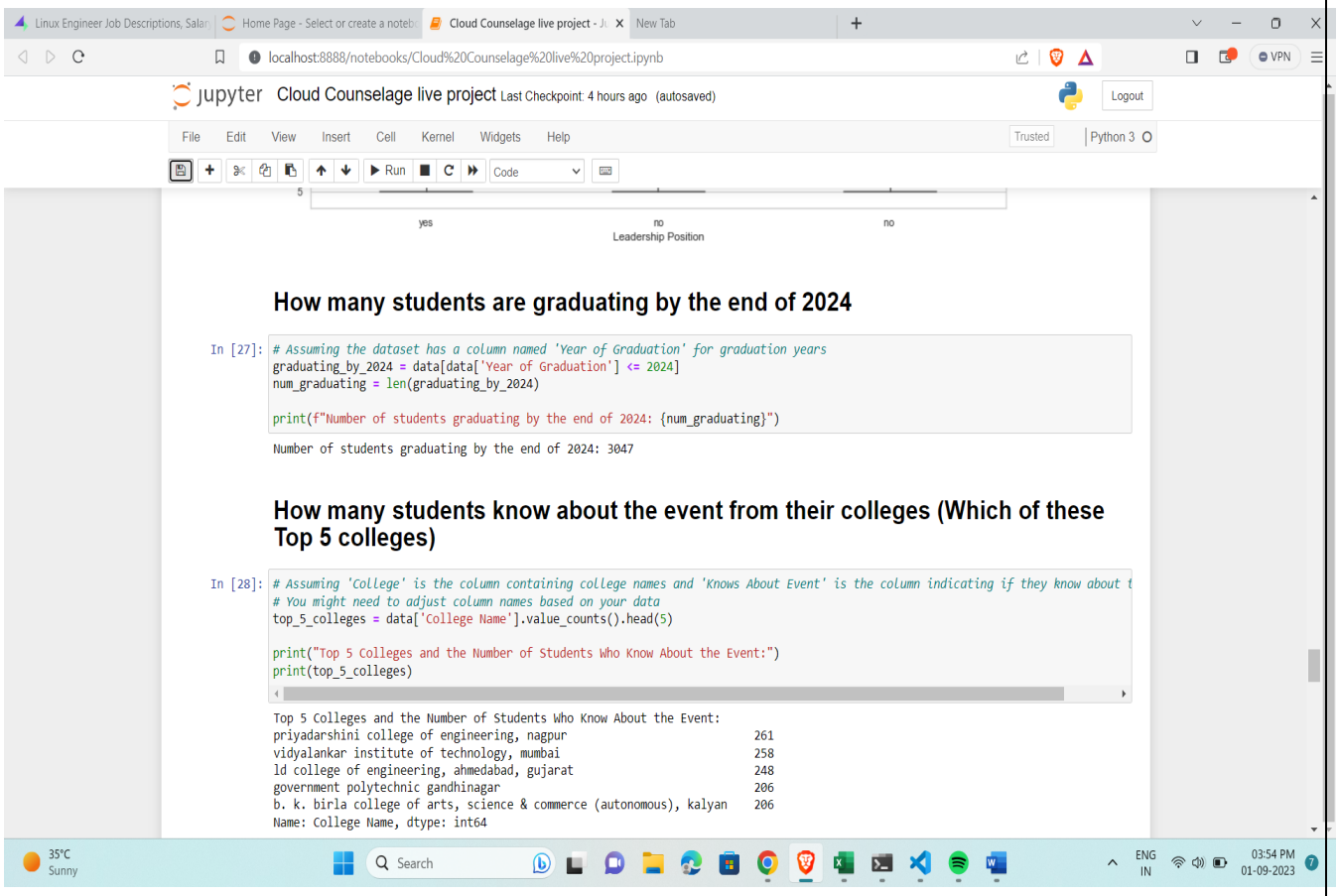
In [25]:
```
# Assuming the dataset has columns 'Events' and 'Attendee Status' for event name and student field
plt.figure(figsize=(12, 6))
sns.countplot(x='Events', hue='Attendee Status', data=data)
plt.title('Event Attendance by Student Fields')
plt.xlabel('Event Name')
plt.ylabel('Number of Students')
plt.xticks(rotation=45)
plt.tight_layout()
plt.legend(title='Student Field')
plt.show()
```

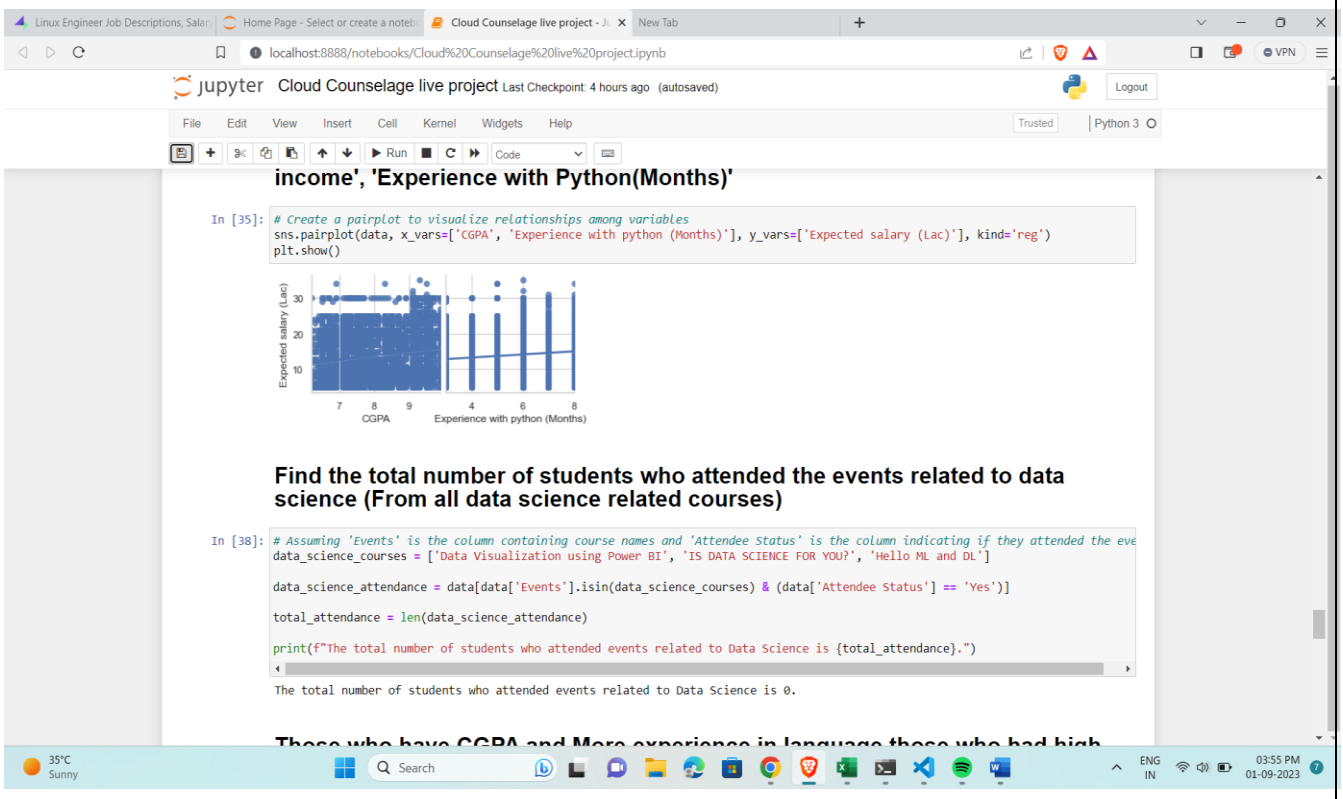


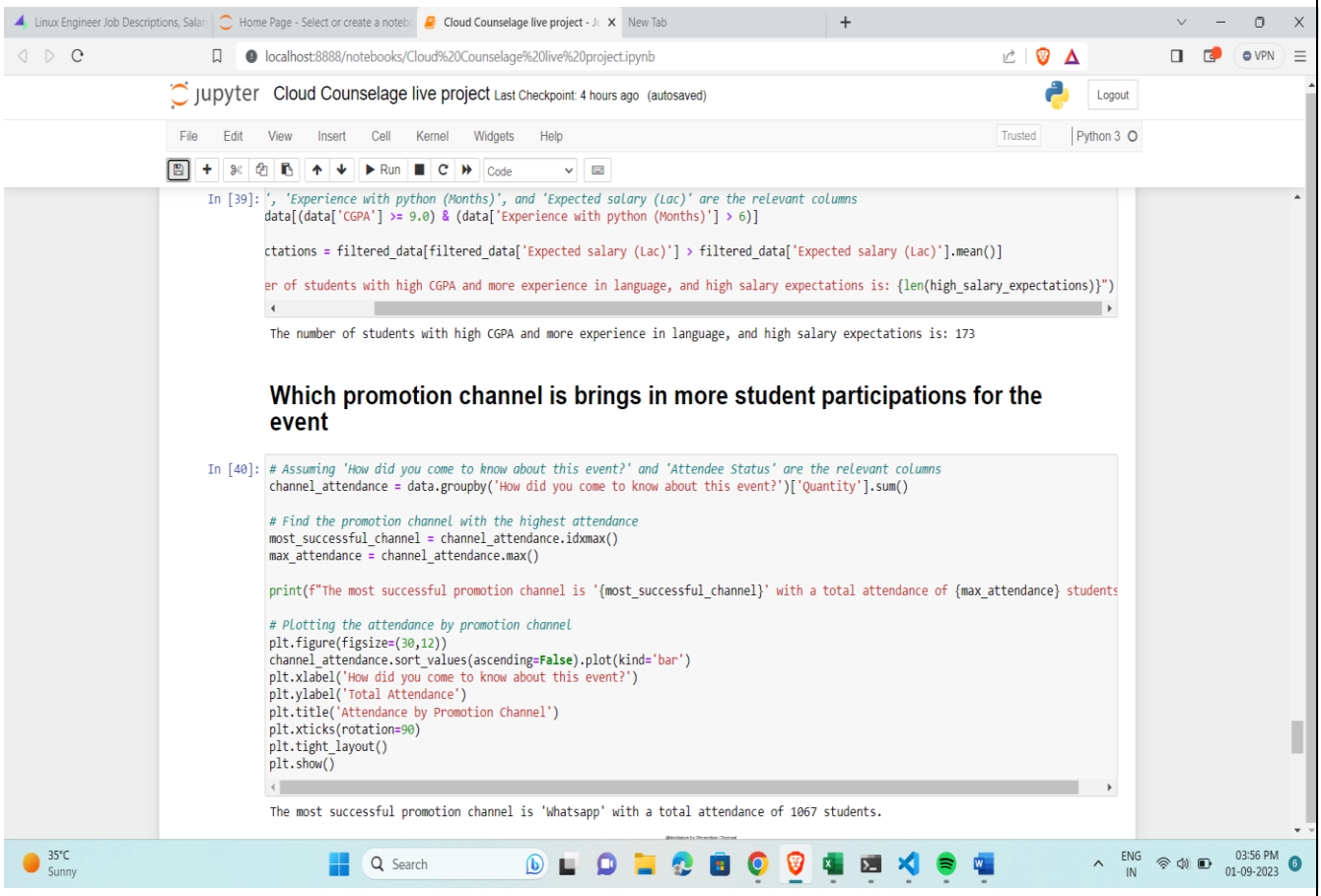
35°C Sunny | Search | ENG IN | 03:52 PM 01-09-2023

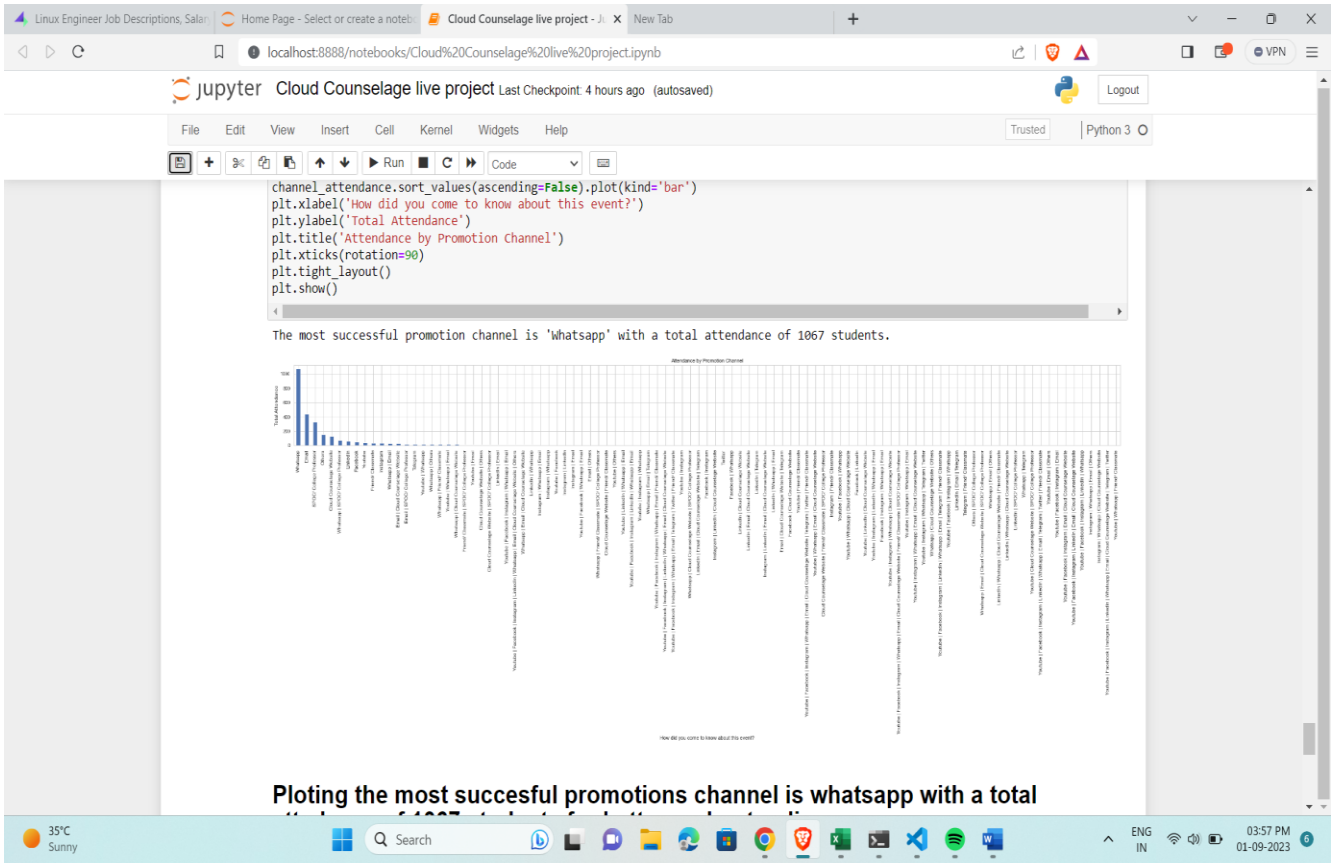

Cloud Counselage live project - Ju
localhost:8888/notebooks/Cloud%20Counselage%20live%20project.ipynb
jupyter Cloud Counselage live project Last Checkpoint: 4 hours ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
plt.xlabel('Leadership Position')
plt.ylabel('GPA')
plt.tight_layout()
plt.show()
plt.figure(figsize=(12, 6))
sns.boxplot(x='Leadership- skills', y='Expected salary (Lac)', data=data)
plt.title('Expected Salary vs. Leadership Position')
plt.xlabel('Leadership Position')
plt.ylabel('Expected Salary')
plt.tight_layout()
plt.show()
GPA vs. Leadership Position
GPA
10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
yes
no
no
Leadership Position
35°C
Sunny
Search
ENG
IN
03:53 PM
01-09-2023

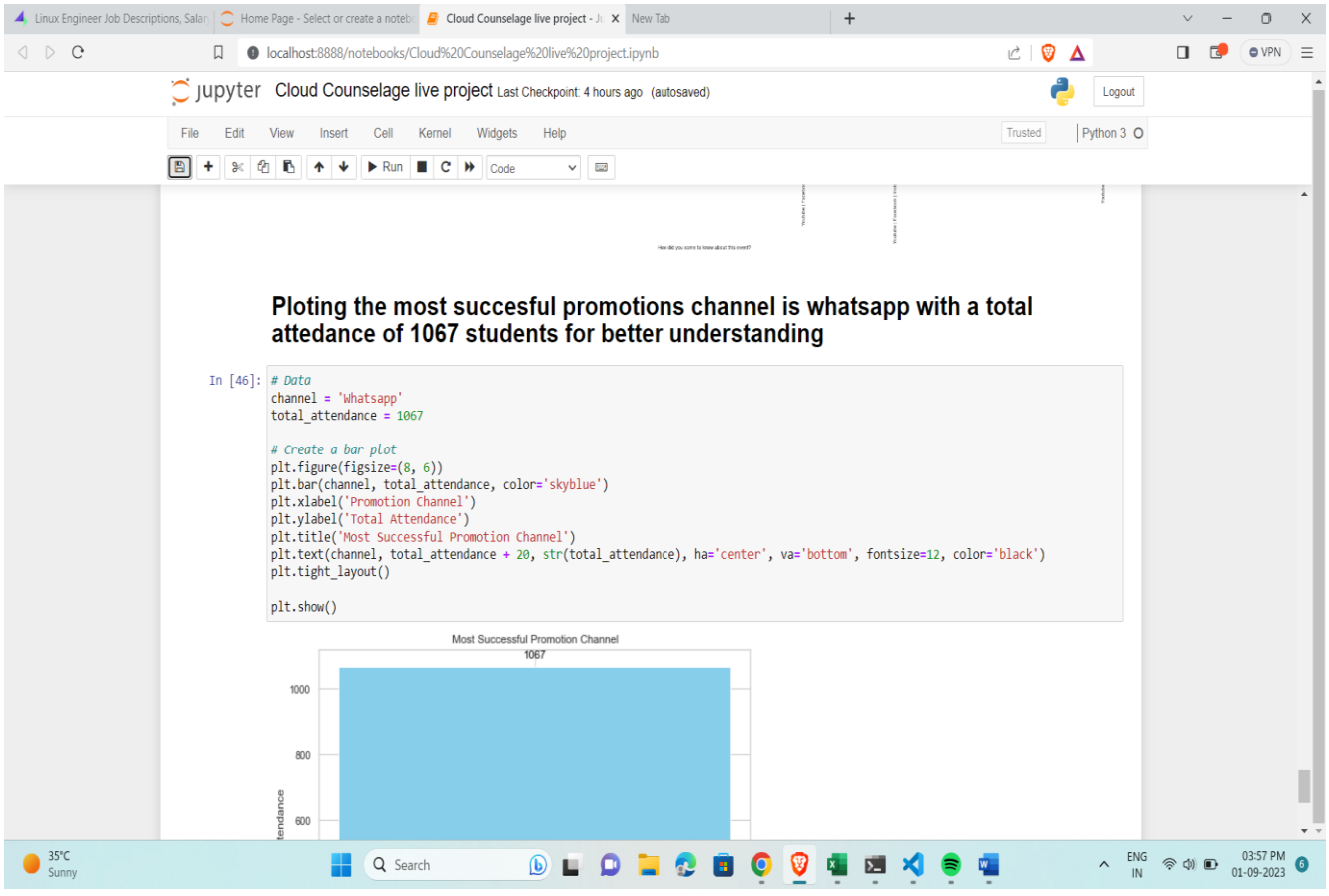

Cloud Counselage live project - Ju
localhost:8888/notebooks/Cloud%20Counselage%20live%20project.ipynb
jupyter Cloud Counselage live project Last Checkpoint: 4 hours ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
7.0
6.5
yes
no
no
Leadership Position
Expected Salary vs. Leadership Position
Expected Salary
35
30
25
20
15
10
5
yes
no
no
Leadership Position
How many students are graduating by the end of 2024
35°C
Sunny
Search
ENG
IN
03:54 PM
01-09-2023

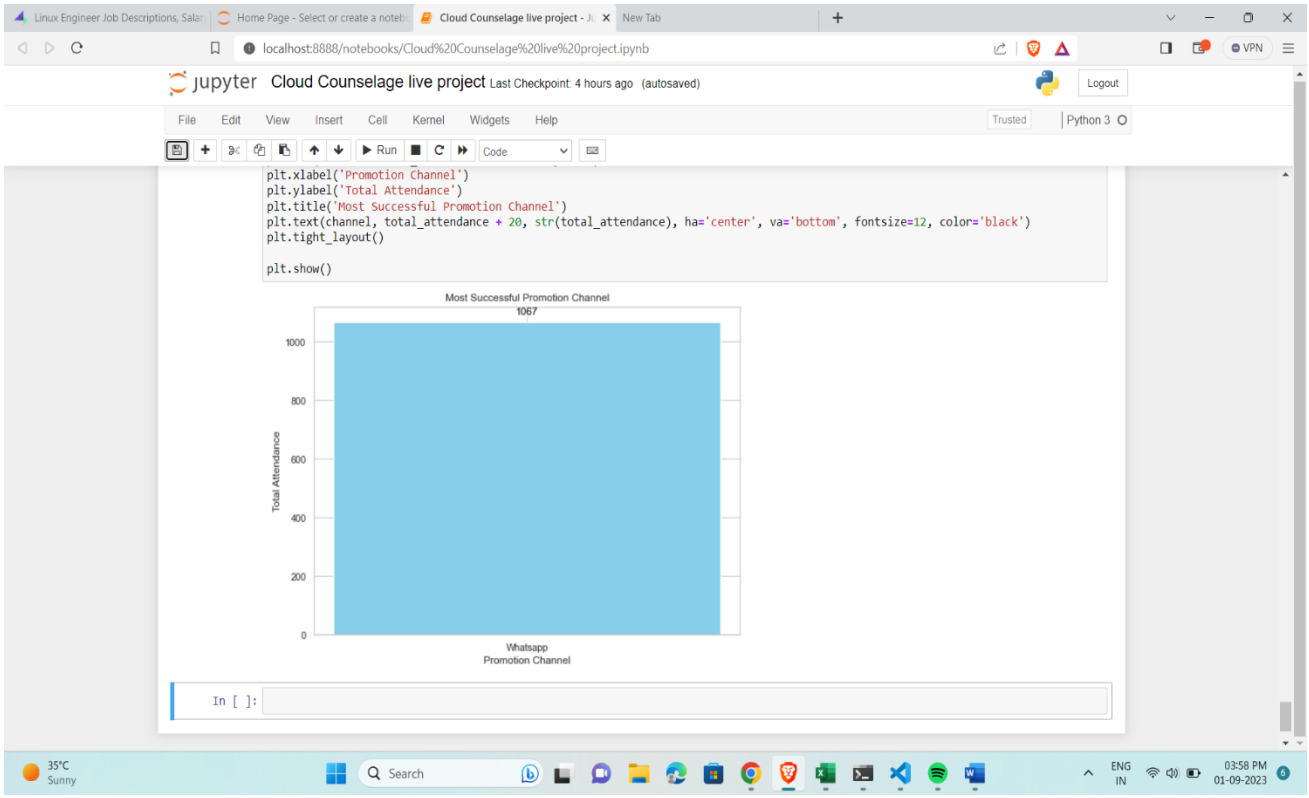

Linux Engineer Job Descriptions, Salar
Home Page - Select or create a notebo
Cloud Counselage live project - J
New Tab
localhost:8888/notebooks/Cloud%20Counselage%20live%20project.ipynb
jupyter Cloud Counselage live project Last Checkpoint: 4 hours ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
5
yes
no
no
Leadership Position
How many students are graduating by the end of 2024
In [27]:
# Assuming the dataset has a column named 'Year of Graduation' for graduation years
graduating_by_2024 = data[data['Year of Graduation'] <= 2024]
num_graduating = len(graduating_by_2024)
print(f"Number of students graduating by the end of 2024: {num_graduating}")
Number of students graduating by the end of 2024: 3047
How many students know about the event from their colleges (Which of these Top 5 colleges)
In [28]:
# Assuming 'College' is the column containing college names and 'Knows About Event' is the column indicating if they know about t
# You might need to adjust column names based on your data
top_5_colleges = data['College Name'].value_counts().head(5)
print("Top 5 Colleges and the Number of Students Who Know About the Event:")
print(top_5_colleges)
Top 5 Colleges and the Number of Students Who Know About the Event:
priyadarshini college of engineering, nagpur 261
vidyalankar institute of technology, mumbai 258
ld college of engineering, ahmedabad, gujarat 248
government polytechnic gandhinagar 206
b. k. birla college of arts, science & commerce (autonomous), kalyan 206
Name: College Name, dtype: int64
35°C
Sunny
Search
ENG
IN
03:54 PM
01-09-2023


Linux Engineer Job Descriptions, Salar
Home Page - Select or create a notebo
Cloud Counselage live project - J
New Tab
localhost:8888/notebooks/Cloud%20Counselage%20live%20project.ipynb
jupyter Cloud Counselage live project Last Checkpoint: 4 hours ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
income', 'Experience with Python(Months)'
In [35]:
# Create a pairplot to visualize relationships among variables
sns.pairplot(data, x_vars=['CGPA', 'Experience with python (Months)'], y_vars=['Expected salary (Lac)'], kind='reg')
plt.show()
Expected salary (Lac)
30
20
10
7
8
9
4
6
8
CGPA
Experience with python (Months)
Find the total number of students who attended the events related to data science (From all data science related courses)
In [38]:
# Assuming 'Events' is the column containing course names and 'Attendee Status' is the column indicating if they attended the eve
data_science_courses = ['Data Visualization using Power BI', 'IS DATA SCIENCE FOR YOU?', 'Hello ML and DL']
data_science_attendance = data[data['Events'].isin(data_science_courses) & (data['Attendee Status'] == 'Yes')]
total_attendance = len(data_science_attendance)
print(f"The total number of students who attended events related to Data Science is {total_attendance}.")
The total number of students who attended events related to Data Science is 0.
35°C
Sunny
Search
ENG
IN
03:55 PM
01-09-2023


Linux Engineer Job Descriptions, Salar
Home Page - Select or create a noteb
Cloud Counselage live project - Ju
New Tab
localhost:8888/notebooks/Cloud%20Counselage%20live%20project.ipynb
jupyter Cloud Counselage live project Last Checkpoint: 4 hours ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
In [39]:
', 'Experience with python (Months)', and 'Expected salary (Lac)' are the relevant columns
data[(data['CGPA'] >= 9.0) & (data['Experience with python (Months)'] > 6)]
ctations = filtered_data[filtered_data['Expected salary (Lac)'] > filtered_data['Expected salary (Lac)'].mean()]
er of students with high CGPA and more experience in language, and high salary expectations is: {len(high_salary_expectations)}")
The number of students with high CGPA and more experience in language, and high salary expectations is: 173
Which promotion channel is brings in more student participations for the event
In [40]:
# Assuming 'How did you come to know about this event?' and 'Attendee Status' are the relevant columns
channel_attendance = data.groupby('How did you come to know about this event?')['Quantity'].sum()
# Find the promotion channel with the highest attendance
most_successful_channel = channel_attendance.idxmax()
max_attendance = channel_attendance.max()
print(f"The most successful promotion channel is '{most_successful_channel}' with a total attendance of {max_attendance} students
# Plotting the attendance by promotion channel
plt.figure(figsize=(30,12))
channel_attendance.sort_values(ascending=False).plot(kind='bar')
plt.xlabel('How did you come to know about this event?')
plt.ylabel('Total Attendance')
plt.title('Attendance by Promotion Channel')
plt.xticks(rotation=90)
plt.tight_layout()
plt.show()
The most successful promotion channel is 'Whatsapp' with a total attendance of 1067 students.
35°C
Sunny
Search
ENG
IN
03:56 PM
01-09-2023


Linux Engineer Job Descriptions, Salar
Home Page - Select or create a noteb
Cloud Counselage live project - J
New Tab
localhost:8888/notebooks/Cloud%20Counselage%20live%20project.ipynb
jupyter Cloud Counselage live project Last Checkpoint: 4 hours ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
channel_attendance.sort_values(ascending=False).plot(kind='bar')
plt.xlabel('How did you come to know about this event?')
plt.ylabel('Total Attendance')
plt.title('Attendance by Promotion Channel')
plt.xticks(rotation=90)
plt.tight_layout()
plt.show()
The most successful promotion channel is 'Whatsapp' with a total attendance of 1067 students.
Ploting the most succesful promotions channel is whatsapp with a total
35°C
Sunny
Search
ENG
IN
03:57 PM
01-09-2023


Linux Engineer Job Descriptions, Salar
Home Page - Select or create a noteb
Cloud Counselage live project - J
New Tab
localhost:8888/notebooks/Cloud%20Counselage%20live%20project.ipynb
jupyter Cloud Counselage live project Last Checkpoint: 4 hours ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
Ploting the most succesful promotions channel is whatsapp with a total attedance of 1067 students for better understanding
In [46]:
# Data
channel = 'Whatsapp'
total_attendance = 1067
# Create a bar plot
plt.figure(figsize=(8, 6))
plt.bar(channel, total_attendance, color='skyblue')
plt.xlabel('Promotion Channel')
plt.ylabel('Total Attendance')
plt.title('Most Successful Promotion Channel')
plt.text(channel, total_attendance + 20, str(total_attendance), ha='center', va='bottom', fontsize=12, color='black')
plt.tight_layout()
plt.show()
Most Successful Promotion Channel
1067
1000
800
600
35°C
Sunny
Search
ENG
IN
03:57 PM
01-09-2023



CONCLUSION

In conclusion, this project provides valuable insights into the complex relationship between academic performance, event participation, career aspirations, and factors influencing the success of student interns. These findings can guide educational institutions and employers in enhancing the intern experience and promoting success among student interns.

RECOMMENDATIONS

Based on our findings, we recommend the following:

1. Encourage students to actively participate in extracurricular events and workshops to enhance their academic performance and career prospects.
2. Develop mentorship and support programs to assist student interns in achieving their career goals.
3. Foster the development of personal attributes such as resilience and adaptability through targeted training and workshops.

FUTURE WORK

Future research can expand upon this project by including a larger and more diverse dataset, conducting longitudinal studies to track student intern success over time, and exploring the impact of various internship programs on career outcomes.

REFERENCES

1. Internship Website: - https://www.industryacademiacommunity.com/courses/Internships
2. Pre-Internship Talk: - https://www.industryacademiacommunity.com/sessions/IAC-Pre-Internship-Talk-64902047aeda092a26b204b2
3. Please refer the following Project Management sessions in case you are facing any issue in delivering the projects:

4. Project Initiation: https://cloudcounselage.graphy.com/sessions/Project-Initiation-64af88a3b308530823e2c19e

5. Project Planning: https://cloudcounselage.graphy.com/sessions/Project-Planning-64af88dcb308530823e2¢19f

6. Project Execution: https://cloudcounselage.graphy.com/sessions/Project-Execution--Monitoring-64b156a8b308530823e2c2fa

7. FAQs for Chatbot: https://www.industryacademiacommunity.com/faqs

---

THANK YOU