BASIC & MODERATE IAC QUESTIONS

BASIC QUESTIONS:

1. How many unique students are included in the datasets ?

CODE:

Import pandas as pd

Import numpy as N

list\_inp = ['10 Pawan', 'Aaditya','Aaftab','Aakanksha','Aakanksha','Aakanksha','Aakansha','Aaliya Ruba','Aakassh','Aamir']

res = N.array(list\_inp)

Unique\_res = N.unique(res)

print(“Unique element of the list using numpy .unique ():\n”)

Print(Unique\_res)

OUTPUT:

Unique elements of the list using numpy.unique():

['10 Pawan' 'Aaditya' 'Aaftab' 'Aakanksha' 'Aakansha' 'Aakassh'

'Aaliya Ruba' 'Aamir']

CONCLUSION:

By using numpy library I introduced the module. From the array of list I found the unique value by using “ Unique\_res = N.unique(res) ” this specific code informs the program that no duplicate value should be considered.

1. What is the average GPA of students?

CODE:

Import pandas as pd

Students = [

{"Name": "10 Pawan", "CGPA":7.3},

{"Name": "Aaditya", "CGPA":7.3},

{"Name": "Aaftab", "CGPA":7.3},

{"Name": "Aakansha", "CGPA":7.2},

{"Name": "Aakansha", "CGPA":9.6},

{"Name": "Aakansha", "CGPA":7.5},

{"Name": "Aakansha", "CGPA":9.9},

{"Name": "Aakassh", "CGPA":7.3},

{"Name": "Aliya Ruba", "CGPA":7.2},

{"Name": "Aamir", "CGPA":8.3}

]

average\_cgpa = sum(students["CGPA"] for students in students) / len(students)

print("Average CGPA:", average\_cgpa)

OUTPUT:

Average CGPA: 7.889999999999999

CONCLUSION:

By using the aggregate function I introduced the module. Whereas, average\_gpa found the average of total CGPA of students.

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

CODE:

import pandas as pd

import matplotlib.pyplot as plt

import numpy as np

students = ['10 Pawan', 'Aaditya', 'Aaftab', 'Aakanksha','Aakansha', 'Aakassh', 'Aaliya Ruba', 'Aamir']

data = [2024,2023,2024,2025,2023,2024,2023,2024]

plt.plot(students,data,markersize=5,markeredgecolor='k',linestyle='solid')

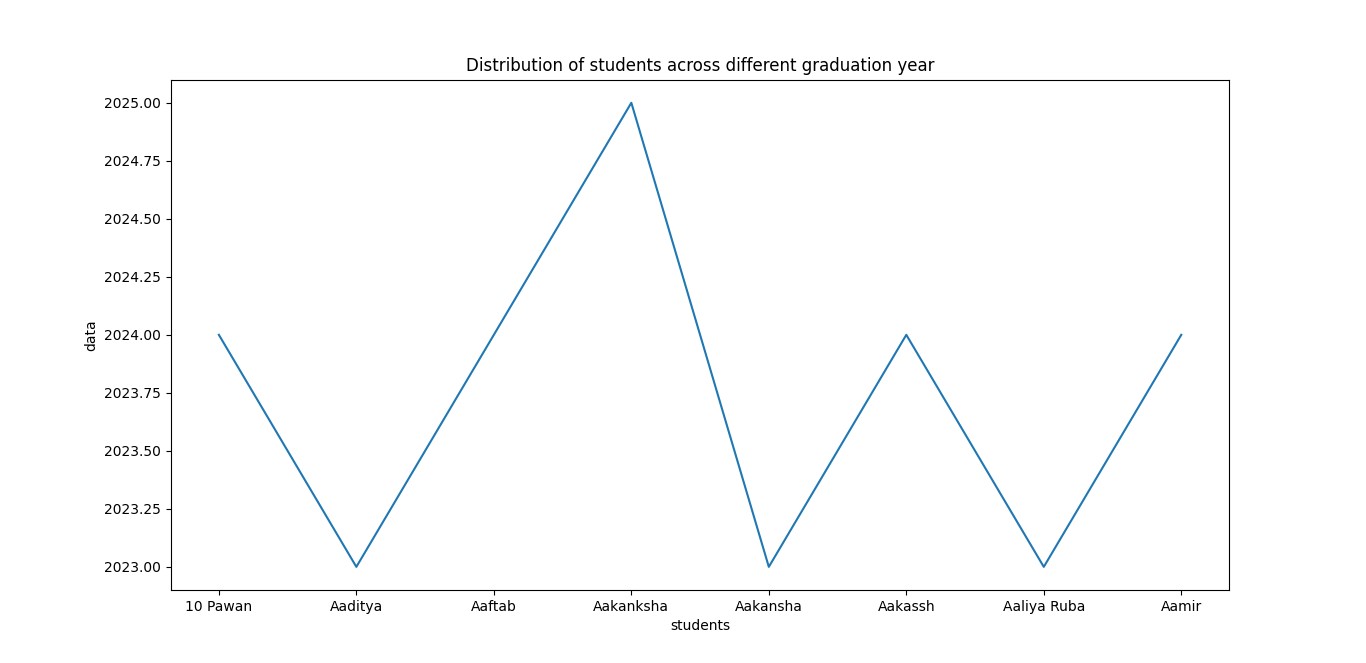
plt.xlabel("students")

plt.ylabel("data")

plt.title("Distribution of students across different graduation year")

plt.show()

OUTPUT:



CONCLUSION:

By using matplotlib.pyplot and numpy library the distribution of students and their different graduation years took place by line graph chart.

4.What is the distribution of student’s experience with python programming?

CODE:

import pandas as pd

import matplotlib.pyplot as plt

students = ["10Pawan","Aaditya","Aaftab","Aakanksha","Aakanksha","Aakanksha","Aakansha","Aakassh","Aaliya Ruba","Aamir"]

Experience = [5,7,7,6,8,4,4,3,4,8]

plt.bar(students,Experience)

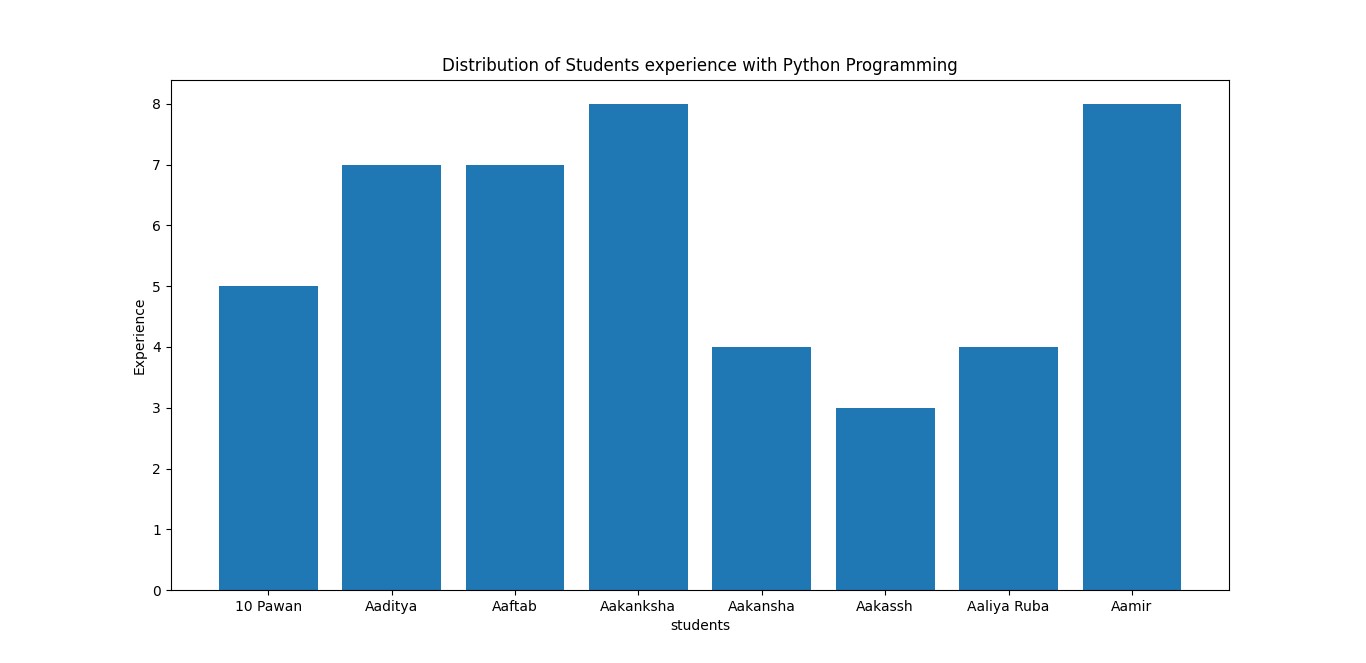
plt.xlabel("students")

plt.ylabel("Experience")

plt.title("Distribution of Students experience with Python Programming")

plt.show()

OUTPUT:



CONCLUSION:

By using Matplotlib.pyplot library the distribution between the students took place that who have how much experience in python by bar chart.

1. What is the average family income of the student?

CODE:

import pandas as pd

income = [2,2,2,2,5,7,2,2,2,2]

average = sum(income)/ len(income)

print("Average of family income: ", (average))

OUTPUT:

Average of family income: 2.8

CONCLUSION:

By using the sum of income of student divided by number of students I found the average family income of the students.

1. How does the GPA vary among different colleges?(Show top 5 results only)

CODE:

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

data = {

'College': ['silicon institute of technology','CHHATTISGARH SWAMI VIVEKANANDA TECHNICAL UNIVERSITY TEACHING DEPARTMENT BHILAI','st xaviers college','wilson college','mit academy of engineering,alandi'],

'GPA': [7.3,7.3,7.3,7.2,9.6]

}

df = pd.DataFrame(data)

avg\_gpa\_by\_college = df.groupby('College')['GPA'].mean().reset\_index()

plt.figure(figsize=(3,3))

plt.bar(avg\_gpa\_of\_college['College'], avg\_gpa\_of\_college['GPA'])

plt.xlabel('College')

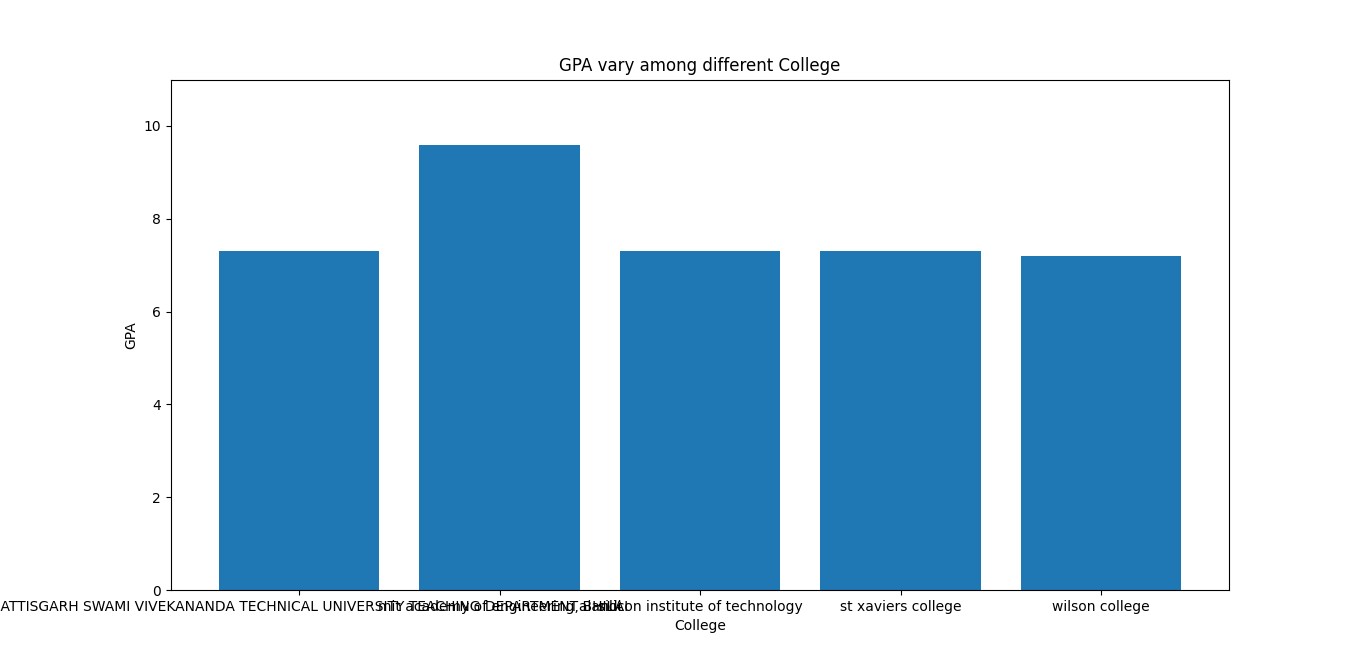
plt.ylabel('GPA')

plt.title('GPA vary among different College')

plt.ylim(0, 11.0)

plt.show()

OUTPUT:



CONCLUSION:

By using Matplotlib.pyplot library the distribution between 5 different type of colleges and it’s GPA took place in python bar chart.

1. What is the average GPA for student from each city?

CODE:

import pandas as pd

data = [

{'student\_name':'10 Pawan', 'city': 'Ghaziabad', 'gpa': 7.3},

{'student\_name': 'Aaditya', 'city': 'Nizambad', 'gpa': 7.3},

{'student\_name': 'Aaftab', 'city': 'Bhsawal', 'gpa': 7.3},

{'student\_name': 'Aakanksha', 'city': 'Santipur', 'gpa': 7.2},

{'student\_name': 'Aakanksha', 'city': 'Jalor', 'gpa': 9.6}

]

df = pd.DataFrame(data)

average\_gpa\_per\_city = df.groupby('city')['gpa'].mean()

print(average\_gpa\_per\_city)

OUTPUT:

Bhsawal 7.3

Ghaziabad 7.3

Jalor 9.6

Nizambad 7.3

Santipur 7.2

CONCLUSION:

By using df.groupby function in city and gpa and finding average of all the list given above .

1. Can we identify any relationship between Family Income and GPA?

CODE:

import pandas as pd

import matplotlib.pyplot as plt

import numpy as np

Info = ['10 Pawan','Aaditya','Aaftab','Aakanksha','Aakanksha1','Aakanksha2','Aakansha','Aakassh','Aaliya Ruba','Aamir']

Income = [2,2,2,2,5,7,2,2,2,2]

GPA = [7.3,7.3,7.3,7.2,9.6,7.5,9.9,7.3,7.2,8.3]

X = np.arange(len(Info))

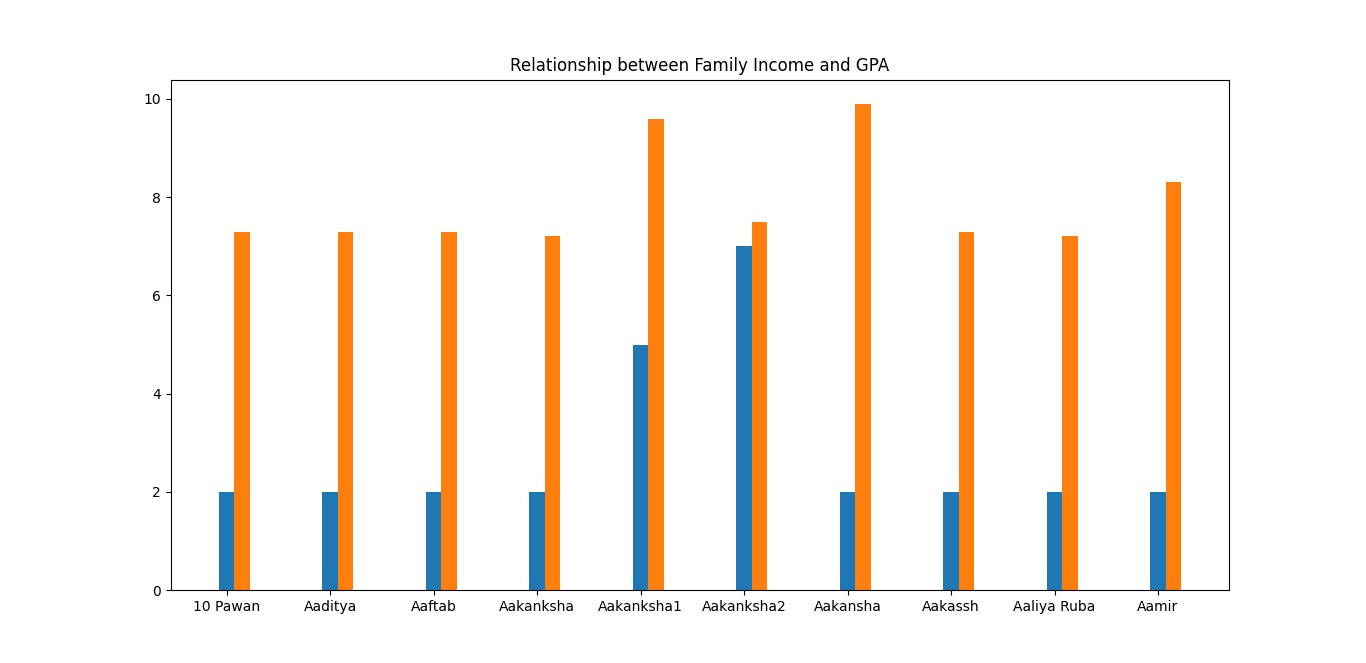
plt.bar(Info,Income,width=.15)

plt.bar(X+0.15,GPA,width=.15)

plt.title('Relationship between Family Income and GPA')

plt.show()

OUTPUT:



CONCLUSION:

By using Matplotlib.pyplot library the distribution of relationship between the Family Income and GPA took place in python bar chart.

MODERATE QUESTIONS:

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

CODE:

import pandas as pd

import matplotlib.pyplot as plt

data = {

'GPA': [7.3,7.3,7.3,7.2,9.6,7.5,9.9,7.3,7.2,8.3],

'FamilyIncome': [2,2,2,2,5,7,2,2,2,2],

'ExperienceMonths': [5,7,7,6,8,4,4,3,4,8],

'ExpectedSalary': [5,8,19,9,21,21,21,13,9,9]

}

df = pd.DataFrame(data)

X= df[['GPA', 'FamilyIncome', 'ExperienceMonths']]

y = df['ExpectedSalary']

plt.scatter(df['ExperienceMonths'], df['ExpectedSalary'], color='Black', label='Data Points')

plt.plot(df['ExperienceMonths'], color='Blue', linewidth=2, label='Regression Line')

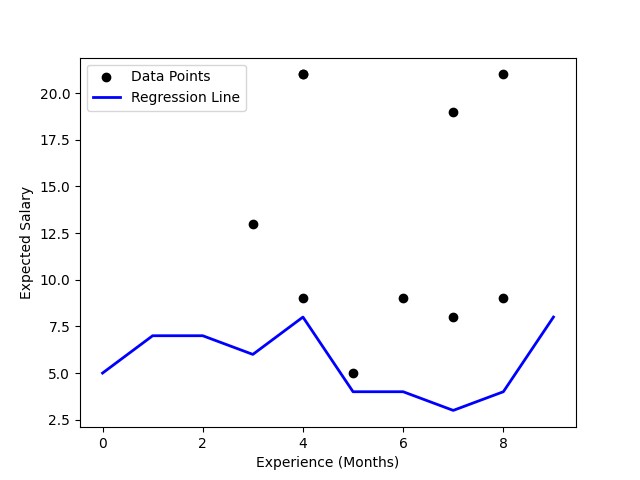
plt.xlabel('Experience (Months)')

plt.ylabel('Expected Salary')

plt.legend()

plt.show()

OUTPUT:



CONCLUSION:

By using matplotlib.pyplot library and with the help of scatter function I have found out the expected salary based on all the factors given above. I have used plt.legend() command for specify the color or mark linked to a specific data range plotted.

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

CODE:

import pandas as pd

import matplotlib.pyplot as plt

data = {

'Event': ['Data Visualization Using Power BI','Product Design & Full Style', 'Art of Resume Building', 'Talk on Skill and Employability Enhancement','Internship Program(IP)Success Conclave'],

'Field\_of\_Study': ['Computer Science', 'Engineering', 'Career Enhancement', 'Career Enhancement', 'Career Enhancement'],

'Number\_of\_Students': [455,842,478,379,797]

}

df = pd.DataFrame(data)

event\_field\_summary = df.groupby('Field\_of\_Study')['Number\_of\_Students']

plt.figure(figsize=(6, 4))

for event in df['Event'].unique():

subset = df[df['Event'] == event]

plt.bar(subset['Field\_of\_Study'], subset['Number\_of\_Students'])

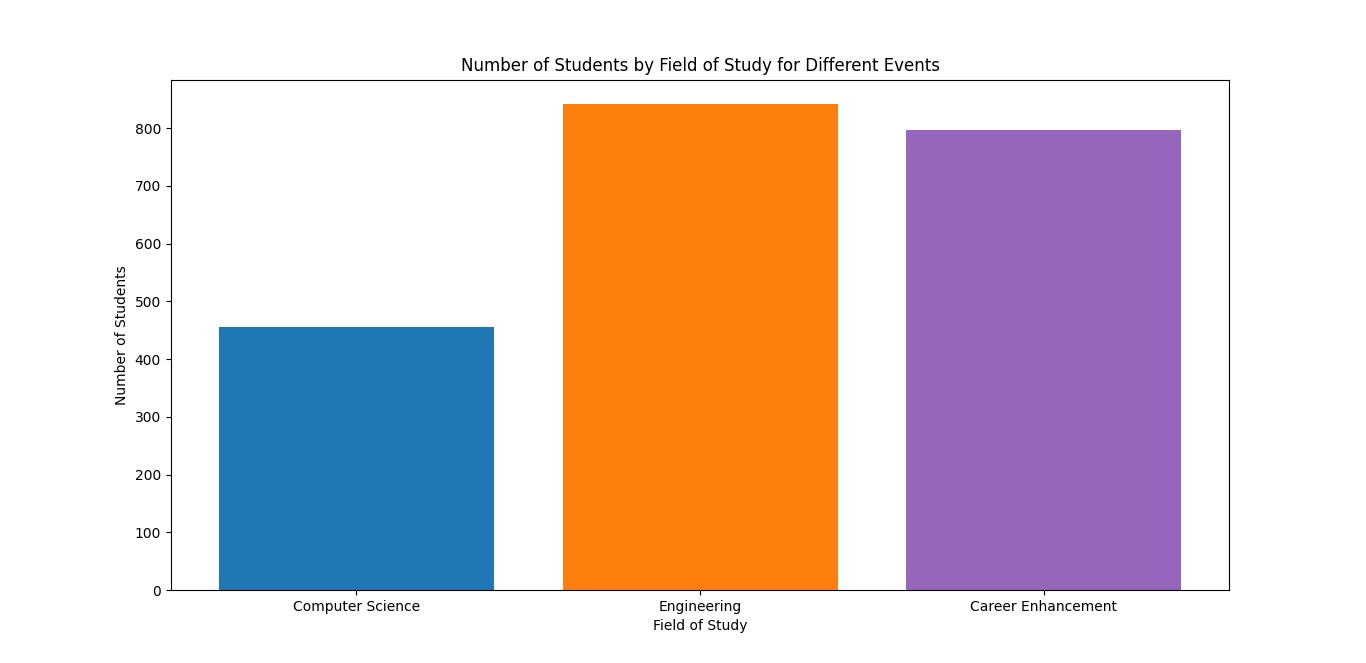
plt.xlabel('Field of Study')

plt.ylabel('Number of Students')

plt .title('Number of Students by Field of Study for Different Events')

plt.show()

OUTPUT:



CONCLUSION:

I found that by using subset under (Number of Students) and (Field of study) most students were categorized of there interested events .

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

CODE:

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

leadership\_skills = np.array([0,1,1,0,0,1,0,0,0,1])

expected\_salary = np.array([5,8,19,9,21,21,21,13,9,9])

r = np.corrcoef(leadership\_skills, expected\_salary)[0, 1]

print("Correlation coefficient (r):", r)

plt.scatter(leadership\_skills, expected\_salary)

plt.xlabel("leadership skills")

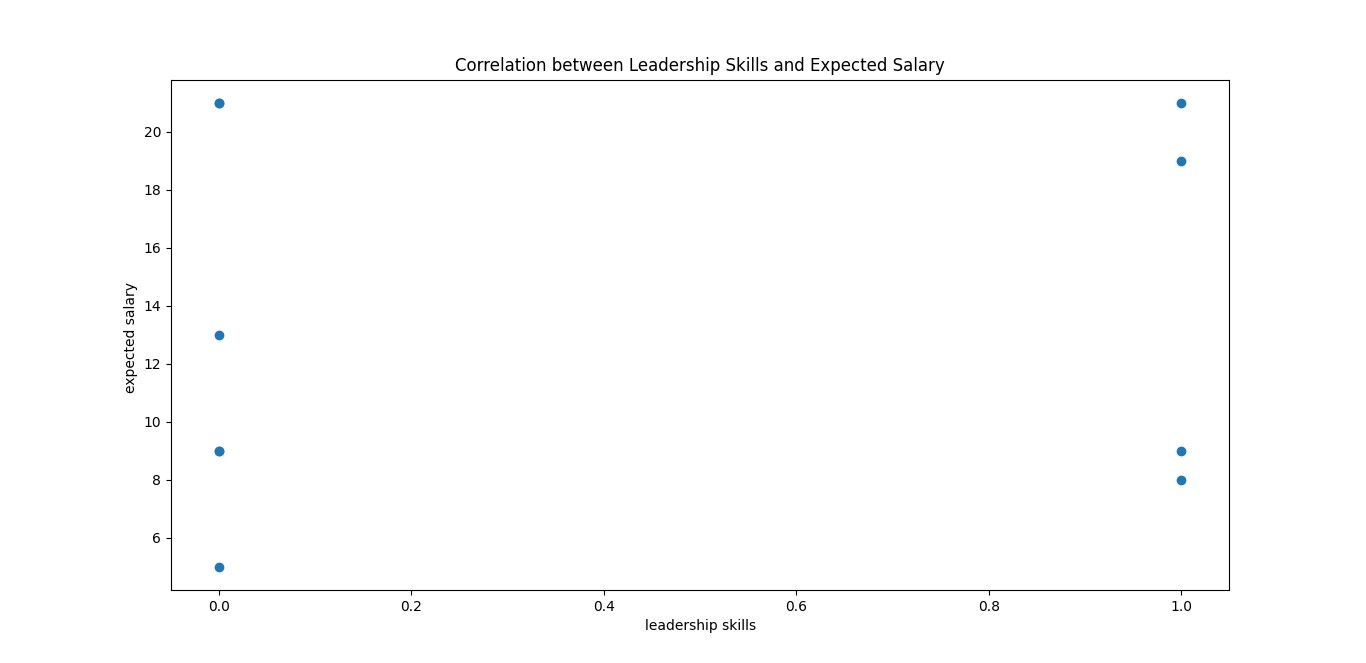
plt.ylabel("expected salary ")

plt.title("Correlation between Leadership Skills and Expected Salary")

plt.show()

OUTPUT:

Correlation coefficient (r): 0.10170952554312156



CONCLUSION:

By using the Scatter Chart I found the correlation between the leadership skills and expected salary were (0) is depicted as [True] and (1) is depicted as [No] -->leadership skill and the expected\_salary is counted in Lakhs. The Dot of the scatter chart points out the expected salary according to 0[True] and 1[No].

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

CODE:

import pandas as pd

graduation\_data = {

2023: 1536,

2024: 1511

}

total\_graduates = sum(graduation\_data.values())

print(f"Total number of students graduating by the end of 2024: {total\_graduates}")

OUTPUT:

Total number of students graduating by the end of 2024: 3047

CONCLUSION:

By using Sum function of python and adding the students of year 2023 and 2024 I found the total graduates of the year 2024.

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

CODE:

import pandas as pd

import matplotlib.pyplot as plt

promotion\_channel = [

'Email',

'Cloud Counselage Website',

'Facebook',

'Instagram',

'Whatsapp',

'Youtube',

'Twitter',

'Linkedin',

'Others'

]

participation\_data = [438,129,48,29,410,37,8,153,49]

plt.bar(promotion\_channel,participation\_data)

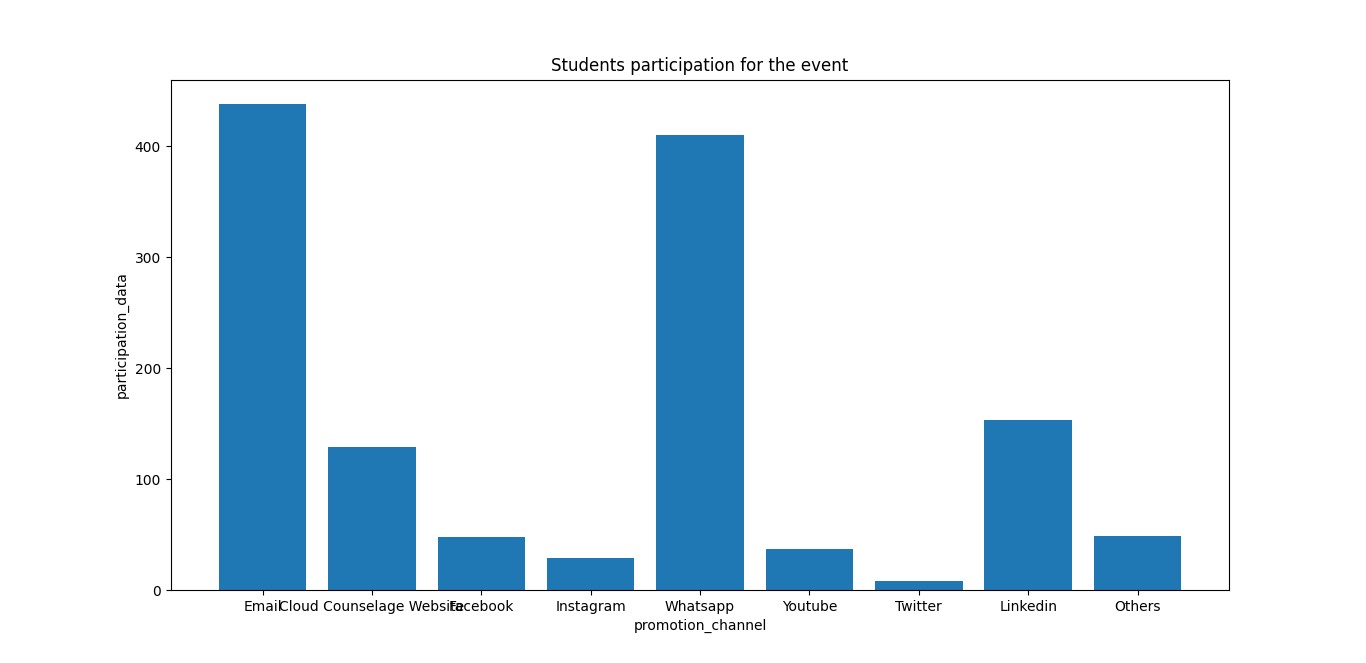
plt.xlabel("promotion\_channel")

plt.ylabel("participation\_data")

plt.title("Students participation for the event")

plt.show()

OUTPUT:



CONCLUSION:

By using matplotlib.pyplot I found that through Email promotion channel most student participated in the event. Whereas, it is represented using bar chart.

16.Find the total number of students who attended the events related to Data Science? (From all Data Science related courses.)

CODE:

import pandas as pd

events = [

{"event\_name": "Is Data Science For You?", "students\_attended": 306},

{"event\_name": "Data Visualization using Power BI", "students\_attended": 455},

]

total\_students = 0

for event in events:

total\_students += event["students\_attended"]

print("Total number of students who attended data science events:", total\_students)

OUTPUT:

Total number of students who attended data science events: 761

CONCLUSION:

By taking total\_students as 0. I have concluded the total students to attend the event. If 1 was placed instead of 0 it would have added 1 to the answer by giving the false answer.

1. Those who have high CGPA and more experience in language those who had high expectations for salary?(Avg)

CODE:

import pandas as pd

data = [

{'CGPA': 9.6, 'LanguageExperience': 8, 'SalaryExpectation': 2100000},

{'CGPA': 9.9, 'LanguageExperience': 4, 'SalaryExpectation': 2100000},

{'CGPA': 9.7, 'LanguageExperience': 6, 'SalaryExpectation': 1900000},

{'CGPA': 9.9, 'LanguageExperience': 4, 'SalaryExpectation': 900000},

{'CGPA': 9.5, 'LanguageExperience': 5, 'SalaryExpectation': 700000}

]

high\_cgpa\_threshold = 9.5

high\_language\_experience\_threshold = 3

filtered\_data = [entry for entry in data if entry['CGPA'] >= high\_cgpa\_threshold and entry['LanguageExperience'] >= high\_language\_experience\_threshold]

if filtered\_data:

average\_salary = sum(entry['SalaryExpectation'] for entry in filtered\_data) / len(filtered\_data)

else:

average\_salary = 0

print(f' Average salary expectation for those with high CGPA and more experience in language: {average\_salary}')

OUTPUT:

Average salary expectation for those with high CGPA and more experience in language: 1540000.0

CONCLUSION:

By specifying high\_cgpa\_thershold as 9.5 and high\_language\_thershold as 3 I have filtered the average salary expectations by calculating the total and dividing the total by number of datas available .

1. How many students know about the event from their colleges? Which of these Top 5 colleges?

CODE:

import pandas as pd

college\_data = {

'A P SHAH INSTITUTE OF TECHNOLOGY': 1,

'GOVERNMENT POLYTECHNIC GANDHINAGAR': 5,

'Wilson College': 9,

'DKTE Societys Textile And Engineering Institute Ichalkaranji': 2,

'Pillai College of Engineering New Panvel': 4,

'LD College of engineering, Ahmedabad, Gujarat': 4,

'St.Francis Institute of Technology': 2,

'S.I.E.S. Graduate School Of Technology, Nerul, Navi Mumbai': 3,

'B. K. Birla College of Arts, Science & Commerce (Autonomous)': 2,

'Don Bosco College of Engineering Fatorda Goa': 1

}

top\_5\_colleges = [

'A P SHAH INSTITUTE OF TECHNOLOGY',

'GOVERNMENT POLYTECHNIC GANDHINAGAR',

'Wilson College',

'DKTE Societys Textile And Engineering Institute Ichalkaranji',

'Pillai College of Engineering New Panvel'

]

total\_students = 0

for college in top\_5\_colleges:

if college in college\_data:

total\_students += college\_data[college]

print(f"Total number of students who know about the event from the top 5 colleges: {total\_students}")

OUTPUT:

Total number of students who know about the event from the top 5 colleges: 21

CONCLUSION:

By taking total\_students as 0. I have concluded the total students who know about the event from their college is 21 . If 1 was placed instead of 0 it would have added 1 to the answer by giving the false answer.