BASIC QUESTIONS : (ATTEMPT ANY 8/9)

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
df = pd.read_excel("/content/Data analyst Data.xlsx")
df.head()
```

	First Name	Email ID	Quantity	Events	Attendee Status	College Name	How did you come to know about this event?	Specify in "Others" (how did you come to know about this event)	Designation	Year of Graduation	City
0	ANIKET	aniket@xyz.com	1	Art of Resume Building	Attending	D Y PATIL INSTITUTE OF MCA AND MANAGEMENT AKUR...	Email	NaN	Students	2024	Patna
1	Dhanshree	dhanshree@xyz.com	1	Art of Resume Building	Attending	AP SHAH INSTITUTE OF TECHNOLOGY	Others	College	Students	2024	Amravati
2	Dhiraj	dhiraj@xyz.com	1	Art of Resume Building	Attending	Don Bosco College of Engineering Fatorda Goa	Email	NaN	Students	2023	Kunrool
3	Pooja	pooja@xyz.com	1	Art of Resume Building	Attending	Pillai College of Engineering New Panvel	Email	NaN	Students	2025	Tirupati
4	Aayush	aayush@xyz.com	1	Art of Resume Building	Attending	St Xavier's College	Instagram | LinkedIn | Cloud Counselage Website	NaN	Students	2024	Vijaywada

Next steps: Generate code with df View recommended plots

```
df.columns
```

```
Index(['First Name', 'Email ID', 'Quantity', 'Events', 'Attendee Status',
       'College Name', 'How did you come to know about this event?',
       'Specify in "Others" (how did you come to know about this event)',
       'Designation', 'Year of Graduation', 'City', 'CGPA',
       'Experience with python (Months)', 'Family Income',
       'Expected salary (Lac)', 'Leadership- skills'],
      dtype='object')
```

```
df.shape
```

```
(4894, 16)
```

1.How Many Unique Students are included in the dataset.?

```
df["First Name"].nunique()
```

```
2324
```

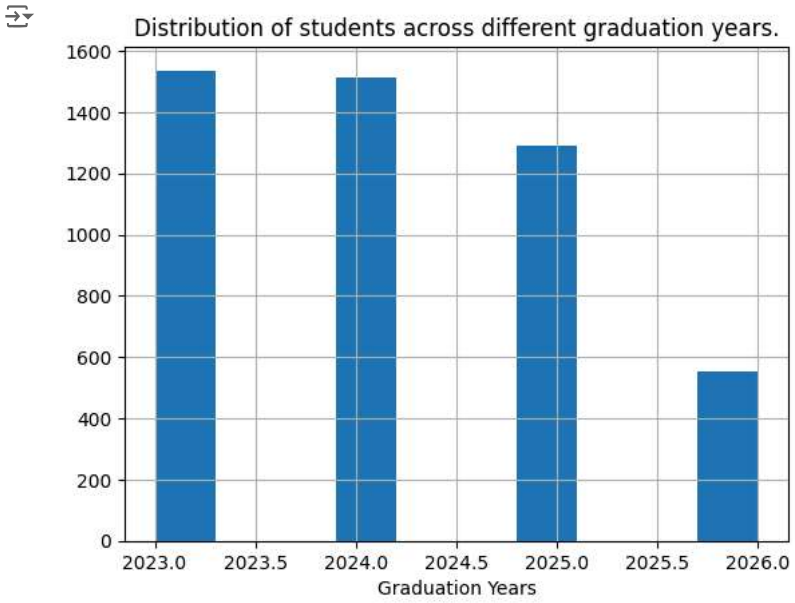
2.What is the average CGPA of the students..?

```
df["CGPA"].mean()
```

```
8.038475684511647
```

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

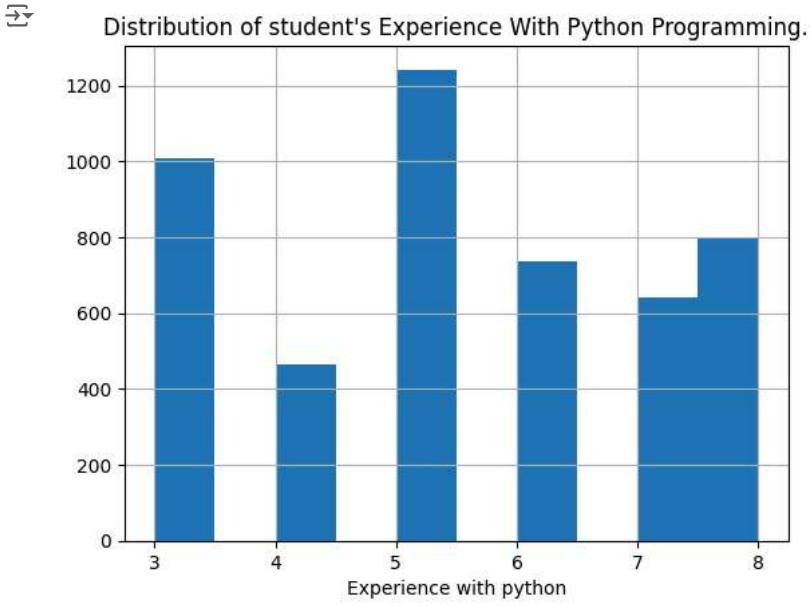
```
plt.title("Distribution of students across different graduation years.")
plt.xlabel("Graduation Years")
plt.hist(x=df["Year of Graduation"])
plt.grid(True)
plt.show()
```



Observation : 2023 have higher distribution than other year of graduation.

4. What is the distribution of student's experience With Python Programming.?

```
plt.title("Distribution of student's Experience With Python Programming.")
plt.xlabel("Experience with python")
plt.hist(x=df["Experience with python (Months)"])
plt.grid(True)
plt.show()
```



Observation : Most of the students rated their Python experience as 5 out of 10.

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

```
new_income = {
    '0-2 Lakh' : 1,
    '2-5 Lakh' : 3.5,
    '5-7 Lakh' : 6,
    '7 Lakh+' : 7
}

df["Average Family Income"] = df["Family Income"].map(new_income)
print("Average Family of the Student is :     ",df["Average Family Income"].mean(),'Lakh')
```

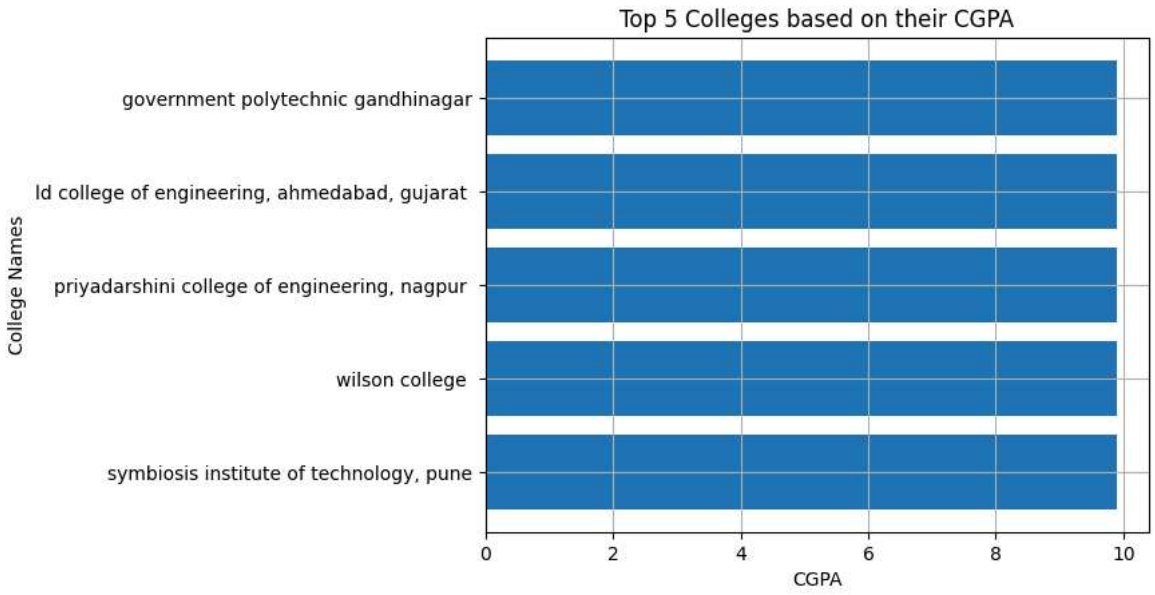
```
Average Family of the Student is :      1.2930118512464241 Lakh
```

6. How does the CGPA Vary amoung different colleges (Show top 5 results only) ?

```
top_colleges = df.sort_values(by='CGPA',ascending=False).head(5)
print(top_colleges["College Name"],top_colleges["CGPA"])


plt.title("Top 5 Colleges based on their CGPA")
plt.xlabel("CGPA")
plt.ylabel("College Names")
plt.barh(top_colleges["College Name"],top_colleges['CGPA'],)
plt.grid(True)
plt.show()
```

```
1003           symbiosis institute of technology, pune
4696                                    wilson college
434      priyadarshini college of engineering, nagpur
3332    ld college of engineering, ahmedabad, gujarat
3324                government polytechnic gandhinagar
Name: College Name, dtype: object 1003    9.9
4696    9.9
434     9.9
3332    9.9
3324    9.9
Name: CGPA, dtype: float64
```



Observation : These are the top 5 colleges with a CGPA of 9.9

8. What is the average CGPA for student from each city..?

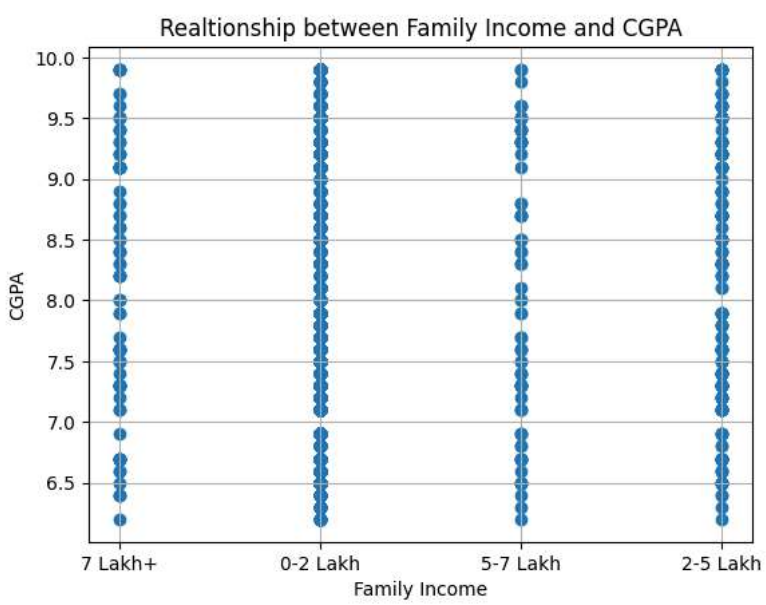
```
df.groupby('City')["CGPA"].mean()
```

```
City
Agartala     7.660714
Agra         8.046429
Ahemdabad    8.190385
```

```
Ajmer        8.284314
Akola        8.021429
               ...
Vidisha      7.738095
Vijaywada    7.986364
Wardha       8.328571
konark       8.071429
kullu        7.878571
Name: CGPA, Length: 177, dtype: float64
```

9. Can We Identify any realtionship between family income and CGPA..?

```
plt.title("Realtionship between Family Income and CGPA")
X=df["Family Income"]
Y=df["CGPA"]
plt.xlabel("Family Income")
plt.ylabel("CGPA")
plt.scatter(x=X,y=Y)
plt.grid(True)
plt.show()
```



Observation: Students whose family income ranges from 0 to 2 lakhs and 2 to 5 lakhs have CGPAs of 8.0 or higher compared to students from other family incomes.

```
Start coding or generate with AI.
```