Data analystic completed in Lab.
(Assignment 1)

Dhanraj Tungar TY-BCA (A)

2022018100094671

Dhanraj Tungar

- # Student Depression Dataset

import pandas as pd

df = pd.read_csv('studentDepression.csv')

df.head()

	id	Gender	Age	City	Profession	Academic Pressure	Work Pressure	CGPA	Study Satisfaction	Job Satisfaction	Sleep Duration	Dietary Habits	Degree	Have you ever had suicidal thoughts ?	Work/Study Hours	Financial Stress
0	2	Male	33.0	Visakhapatnam	Student	5.0	0.0	8.97	2.0	0.0	5-6 hours	Healthy	B.Pharm	Yes	3.0	1.0
1	8	Female	24.0	Bangalore	Student	2.0	0.0	5.90	5.0	0.0	5-6 hours	Moderate	BSc	No	3.0	2.0
2	26	Male	31.0	Srinagar	Student	3.0	0.0	7.03	5.0	0.0	Less than 5 hours	Healthy	BA	No	9.0	1.0
3	30	Female	28.0	Varanasi	Student	3.0	0.0	5.59	2.0	0.0	7-8 hours	Moderate	BCA	Yes	4.0	5.0
4	32	Female	25.0	Jaipur	Student	4.0	0.0	8.13	3.0	0.0	5-6 hours	Moderate	M.Tech	Yes	1.0	1.0

df.tail()

	id	Gender	Age	City	Profession	Academic Pressure	Work Pressure	CGPA	Study Satisfaction	Job Satisfaction	Sleep Duration	Dietary Habits	Degree	Have you ever had suicidal thoughts ?	Work/Study Hours	Financia Stres
27896	140685	Female	27.0	Surat	Student	5.0	0.0	5.75	5.0	0.0	5-6 hours	Unhealthy	Class 12	Yes	7.0	1.
27897	140686	Male	27.0	Ludhiana	Student	2.0	0.0	9.40	3.0	0.0	Less than 5 hours	Healthy	MSc	No	0.0	3.
27898	140689	Male	31.0	Faridabad	Student	3.0	0.0	6.61	4.0	0.0	5-6 hours	Unhealthy	MD	No	12.0	2.
27899	140690	Female	18.0	Ludhiana	Student	5.0	0.0	6.88	2.0	0.0	Less than 5 hours	Healthy	Class 12	Yes	10.0	5.
27900	140699	Male	27.0	Patna	Student	4.0	0.0	9.24	1.0	0.0	Less than 5 hours	Healthy	BCA	Yes	2.0	3.

```
df.shape
```

```
(27901, 18)
```

```
df.describe()
```

	id	Age	Academic Pressure	Work Pressure	CGPA	Study Satisfaction	Job Satisfaction	Work/Study Hours	Financial Stress	Depression
count	27901.000000	27901.000000	27901.000000	27901.000000	27901.000000	27901.000000	27901.000000	27901.000000	27898.000000	27901.000000
mean	70442.149421	25.822300	3.141214	0.000430	7.656104	2.943837	0.000681	7.156984	3.139867	0.585499
std	40641.175216	4.905687	1.381465	0.043992	1.470707	1.361148	0.044394	3.707642	1.437347	0.492645
min	2.000000	18.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000
25%	35039.000000	21.000000	2.000000	0.000000	6.290000	2.000000	0.000000	4.000000	2.000000	0.000000
50%	70684.000000	25.000000	3.000000	0.000000	7.770000	3.000000	0.000000	8.000000	3.000000	1.000000
75%	105818.000000	30.000000	4.000000	0.000000	8.920000	4.000000	0.000000	10.000000	4.000000	1.000000
max	140699.000000	59.000000	5.000000	5.000000	10.000000	5.000000	4.000000	12.000000	5.000000	1.000000

```
df[['Age','CGPA']]
```

	Age	CGPA
0	33.0	8.97
1	24.0	5.90
2	31.0	7.03
3	28.0	5.59
4	25.0	8.13
...	...	...
27896	27.0	5.75
27897	27.0	9.40
27898	31.0	6.61
27899	18.0	6.88
27900	27.0	9.24

27901 rows × 2 columns

Dhanraj Tungar

```
df.loc[1:5]
```

	id	Gender	Age	City	Profession	Academic Pressure	Work Pressure	CGPA	Study Satisfaction	Job Satisfaction	Sleep Duration	Dietary Habits	Degree	Have you ever had suicidal thoughts ?	Work/Study Hours	Financial Stress	Famil Histor o Menta Illnes
1	8	Female	24.0	Bangalore	Student	2.0	0.0	5.90	5.0	0.0	5-6 hours	Moderate	BSc	No	3.0	2.0	Ye
2	26	Male	31.0	Srinagar	Student	3.0	0.0	7.03	5.0	0.0	Less than 5 hours	Healthy	BA	No	9.0	1.0	Ye
3	30	Female	28.0	Varanasi	Student	3.0	0.0	5.59	2.0	0.0	7-8 hours	Moderate	BCA	Yes	4.0	5.0	Ye
4	32	Female	25.0	Jaipur	Student	4.0	0.0	8.13	3.0	0.0	5-6 hours	Moderate	M.Tech	Yes	1.0	1.0	N
5	33	Male	29.0	Pune	Student	2.0	0.0	5.70	3.0	0.0	Less than 5 hours	Healthy	PhD	No	4.0	1.0	N

```
df.iloc[1:5]
```

	id	Gender	Age	City	Profession	Academic Pressure	Work Pressure	CGPA	Study Satisfaction	Job Satisfaction	Sleep Duration	Dietary Habits	Degree	Have you ever had suicidal thoughts ?	Work/Study Hours	Financial Stress	Famil Histor o Menta Illnes
1	8	Female	24.0	Bangalore	Student	2.0	0.0	5.90	5.0	0.0	5-6 hours	Moderate	BSc	No	3.0	2.0	Ye
2	26	Male	31.0	Srinagar	Student	3.0	0.0	7.03	5.0	0.0	Less than 5 hours	Healthy	BA	No	9.0	1.0	Ye
3	30	Female	28.0	Varanasi	Student	3.0	0.0	5.59	2.0	0.0	7-8 hours	Moderate	BCA	Yes	4.0	5.0	Ye
4	32	Female	25.0	Jaipur	Student	4.0	0.0	8.13	3.0	0.0	5-6 hours	Moderate	M.Tech	Yes	1.0	1.0	N

df[df['Age']>30]

:

	id	Gender	Age	City	Profession	Academic Pressure	Work Pressure	CGPA	Study Satisfaction	Job Satisfaction	Sleep Duration	Dietary Habits	Degree	Have you ever had suicidal thoughts ?	Work/Study Hours	Fi
0	2	Male	33.0	Visakhapatnam	Student	5.0	0.0	8.97	2.0	0.0	5-6 hours	Healthy	B.Pharm	Yes	3.0	
2	26	Male	31.0	Srinagar	Student	3.0	0.0	7.03	5.0	0.0	Less than 5 hours	Healthy	BA	No	9.0	
9	62	Male	31.0	Nashik	Student	2.0	0.0	8.38	3.0	0.0	Less than 5 hours	Moderate	LLB	Yes	2.0	
11	91	Male	33.0	Vadodara	Student	3.0	0.0	7.03	4.0	0.0	Less than 5 hours	Healthy	BE	Yes	10.0	
26	186	Male	31.0	Ahmedabad	Student	2.0	0.0	6.08	5.0	0.0	7-8 hours	Moderate	LLB	Yes	3.0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
27876	140536	Male	33.0	Nagpur	Student	1.0	0.0	7.39	4.0	0.0	7-8 hours	Unhealthy	BHM	No	12.0	
27887	140624	Male	32.0	Rajkot	Student	4.0	0.0	9.19	1.0	0.0	5-6 hours	Healthy	MSc	No	7.0	
27889	140631	Male	33.0	Ahmedabad	Student	1.0	0.0	5.70	2.0	0.0	Less than 5 hours	Healthy	MCA	Yes	2.0	
27895	140684	Male	31.0	Lucknow	Student	2.0	0.0	7.27	5.0	0.0	7-8 hours	Moderate	B.Com	Yes	6.0	
27898	140689	Male	31.0	Faridabad	Student	3.0	0.0	6.61	4.0	0.0	5-6 hours	Unhealthy	MD	No	12.0	

6099 rows × 18 columns

```
df.count()
```

```
id                                       27901
Gender                                   27901
Age                                      27901
City                                     27901
Profession                               27901
Academic Pressure                        27901
Work Pressure                            27901
CGPA                                     27901
Study Satisfaction                       27901
Job Satisfaction                         27901
Sleep Duration                           27901
Dietary Habits                           27901
Degree                                   27901
Have you ever had suicidal thoughts ?    27901
Work/Study Hours                         27901
Financial Stress                         27898
Family History of Mental Illness         27901
Depression                               27901
dtype: int64
```

```
df[df['Age']>30].count()
```

```
id                                       6099
Gender                                   6099
Age                                      6099
City                                     6099
Profession                               6099
Academic Pressure                        6099
Work Pressure                            6099
CGPA                                     6099
Study Satisfaction                       6099
Job Satisfaction                         6099
Sleep Duration                           6099
Dietary Habits                           6099
Degree                                   6099
Have you ever had suicidal thoughts ?    6099
Work/Study Hours                         6099
Financial Stress                         6098
Family History of Mental Illness         6099
Depression                               6099
dtype: int64
```

Dhanraj Tungar

df.query('Age > 30')

	id	Gender	Age	City	Profession	Academic Pressure	Work Pressure	CGPA	Study Satisfaction	Job Satisfaction	Sleep Duration	Dietary Habits	Degree	Have you ever had suicidal thoughts ?	Work/Study Hours	Fi
0	2	Male	33.0	Visakhapatnam	Student	5.0	0.0	8.97	2.0	0.0	5-6 hours	Healthy	B.Pharm	Yes	3.0	
2	26	Male	31.0	Srinagar	Student	3.0	0.0	7.03	5.0	0.0	Less than 5 hours	Healthy	BA	No	9.0	
9	62	Male	31.0	Nashik	Student	2.0	0.0	8.38	3.0	0.0	Less than 5 hours	Moderate	LLB	Yes	2.0	
11	91	Male	33.0	Vadodara	Student	3.0	0.0	7.03	4.0	0.0	Less than 5 hours	Healthy	BE	Yes	10.0	
26	186	Male	31.0	Ahmedabad	Student	2.0	0.0	6.08	5.0	0.0	7-8 hours	Moderate	LLB	Yes	3.0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
27876	140536	Male	33.0	Nagpur	Student	1.0	0.0	7.39	4.0	0.0	7-8 hours	Unhealthy	BHM	No	12.0	
27887	140624	Male	32.0	Rajkot	Student	4.0	0.0	9.19	1.0	0.0	5-6 hours	Healthy	MSc	No	7.0	
27889	140631	Male	33.0	Ahmedabad	Student	1.0	0.0	5.70	2.0	0.0	Less than 5 hours	Healthy	MCA	Yes	2.0	
27895	140684	Male	31.0	Lucknow	Student	2.0	0.0	7.27	5.0	0.0	7-8 hours	Moderate	B.Com	Yes	6.0	
27898	140689	Male	31.0	Faridabad	Student	3.0	0.0	6.61	4.0	0.0	5-6 hours	Unhealthy	MD	No	12.0	

6099 rows × 18 columns

```
df.isnull()
```

	id	Gender	Age	City	Profession	Academic Pressure	Work Pressure	CGPA	Study Satisfaction	Job Satisfaction	Sleep Duration	Dietary Habits	Degree	Have you ever had suicidal thoughts ?	Work/Study Hours	Financial Stress	Fam Histo Men Illne
0	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	Fal
1	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	Fal
2	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	Fal
3	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	Fal
4	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	Fal
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
27896	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	Fal
27897	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	Fal
27898	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	Fal
27899	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	Fal
27900	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	Fal

27901 rows × 18 columns

```
df.isnull().sum()
```

```
id                                      0
Gender                                  0
Age                                     0
City                                    0
Profession                              0
Academic Pressure                       0
Work Pressure                           0
CGPA                                    0
Study Satisfaction                      0
Job Satisfaction                        0
Sleep Duration                          0
Dietary Habits                          0
Degree                                  0
Have you ever had suicidal thoughts ?   0
Work/Study Hours                        0
Financial Stress                        3
Family History of Mental Illness        0
Depression                              0
dtype: int64
```

```
#fill missing values with zero
df1 = df.fillna(0)
```

#check the changes count

df1.isnull().sum()

```
id                                        0
Gender                                    0
Age                                       0
City                                      0
Profession                                0
Academic Pressure                         0
Work Pressure                             0
CGPA                                      0
Study Satisfaction                        0
Job Satisfaction                          0
Sleep Duration                            0
Dietary Habits                            0
Degree                                    0
Have you ever had suicidal thoughts ?     0
Work/Study Hours                          0
Financial Stress                          0
Family History of Mental Illness          0
Depression                                0
dtype: int64
```

df.dtypes

```
id                                      int64
Gender                                 object
Age                                   float64
City                                   object
Profession                             object
Academic Pressure                     float64
Work Pressure                         float64
CGPA                                  float64
Study Satisfaction                    float64
Job Satisfaction                      float64
Sleep Duration                         object
Dietary Habits                         object
Degree                                 object
Have you ever had suicidal thoughts ?  object
Work/Study Hours                      float64
Financial Stress                      float64
Family History of Mental Illness       object
Depression                              int64
dtype: object
```

Dhanraj Tungar

df.sort_values(by='Age', ascending=False)

	id	Gender	Age	City	Profession	Academic Pressure	Work Pressure	CGPA	Study Satisfaction	Job Satisfaction	Sleep Duration	Dietary Habits	Degree	Have you ever had suicidal thoughts ?	Work/Study Hours	Financia Stres
9238	46602	Male	59.0	Nashik	Student	1.0	0.0	8.14	1.0	0.0	5-6 hours	Unhealthy	PhD	Yes	10.0	4.(
2909	14768	Female	58.0	Chennai	Student	4.0	0.0	8.58	1.0	0.0	7-8 hours	Healthy	Class 12	No	4.0	4.(
14819	74887	Female	56.0	Ludhiana	Student	3.0	0.0	7.94	5.0	0.0	5-6 hours	Unhealthy	BSc	No	1.0	5.(
13499	68441	Male	54.0	Agra	Student	5.0	0.0	9.60	2.0	0.0	More than 8 hours	Unhealthy	B.Ed	Yes	9.0	3.(
4386	22004	Female	51.0	Bhopal	Student	2.0	0.0	8.26	3.0	0.0	Less than 5 hours	Moderate	MBBS	Yes	5.0	5.(
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	..
15856	80171	Male	18.0	Jaipur	Student	1.0	0.0	8.98	5.0	0.0	5-6 hours	Unhealthy	Class 12	No	10.0	3.(
15870	80239	Female	18.0	Lucknow	Student	3.0	0.0	5.41	3.0	0.0	More than 8 hours	Moderate	Class 12	Yes	11.0	5.(
15871	80246	Female	18.0	Indore	Student	5.0	0.0	9.44	1.0	0.0	More than 8 hours	Unhealthy	Class 12	Yes	7.0	5.(
15873	80270	Male	18.0	Rajkot	Student	4.0	0.0	7.77	3.0	0.0	Less than 5 hours	Moderate	Class 12	Yes	6.0	1.(
22537	113642	Female	18.0	Ludhiana	Student	4.0	0.0	5.39	3.0	0.0	7-8 hours	Unhealthy	Class 12	Yes	7.0	2.(

df.sort_values(by='Age', ascending=False).head()

	id	Gender	Age	City	Profession	Academic Pressure	Work Pressure	CGPA	Study Satisfaction	Job Satisfaction	Sleep Duration	Dietary Habits	Degree	Have you ever had suicidal thoughts ?	Work/Study Hours	Financial Stress
9238	46602	Male	59.0	Nashik	Student	1.0	0.0	8.14	1.0	0.0	5-6 hours	Unhealthy	PhD	Yes	10.0	4.0
2909	14768	Female	58.0	Chennai	Student	4.0	0.0	8.58	1.0	0.0	7-8 hours	Healthy	Class 12	No	4.0	4.0
14819	74887	Female	56.0	Ludhiana	Student	3.0	0.0	7.94	5.0	0.0	5-6 hours	Unhealthy	BSc	No	1.0	5.0
13499	68441	Male	54.0	Agra	Student	5.0	0.0	9.60	2.0	0.0	More than 8 hours	Unhealthy	B.Ed	Yes	9.0	3.0
4386	22004	Female	51.0	Bhopal	Student	2.0	0.0	8.26	3.0	0.0	Less than 5 hours	Moderate	MBBS	Yes	5.0	5.0

df.sort_values(by='Age', ascending=False).head(10)

Dhanraj Tungar

	id	Gender	Age	City	Profession	Academic Pressure	Work Pressure	CGPA	Study Satisfaction	Job Satisfaction	Sleep Duration	Dietary Habits	Degree	Have you ever had suicidal thoughts ?	Work/Study Hours	Financi Stre
9238	46602	Male	59.0	Nashik	Student	1.0	0.0	8.14	1.0	0.0	5-6 hours	Unhealthy	PhD	Yes	10.0	4
2909	14768	Female	58.0	Chennai	Student	4.0	0.0	8.58	1.0	0.0	7-8 hours	Healthy	Class 12	No	4.0	4
14819	74887	Female	56.0	Ludhiana	Student	3.0	0.0	7.94	5.0	0.0	5-6 hours	Unhealthy	BSc	No	1.0	5
13499	68441	Male	54.0	Agra	Student	5.0	0.0	9.60	2.0	0.0	More than 8 hours	Unhealthy	B.Ed	Yes	9.0	3
4386	22004	Female	51.0	Bhopal	Student	2.0	0.0	8.26	3.0	0.0	Less than 5 hours	Moderate	MBBS	Yes	5.0	5
3436	17213	Male	49.0	Nagpur	Student	3.0	0.0	6.47	3.0	0.0	5-6 hours	Moderate	Class 12	Yes	2.0	1
25202	126970	Female	48.0	Rajkot	Student	3.0	0.0	5.32	4.0	0.0	7-8 hours	Unhealthy	Class 12	No	1.0	4
13618	69032	Female	48.0	Ghaziabad	Student	3.0	0.0	7.48	5.0	0.0	More than 8 hours	Healthy	LLB	No	2.0	4
27334	137827	Female	48.0	Surat	Student	1.0	0.0	8.92	4.0	0.0	7-8 hours	Moderate	MD	Yes	12.0	4
9006	45393	Male	46.0	Nashik	Student	1.0	0.0	6.10	3.0	0.0	7-8 hours	Unhealthy	PhD	No	4.0	2

```
df[['CGPA']].min()
```

```
CGPA    0.0
dtype: float64
```

```
df[['CGPA']].max()
```

```
CGPA    10.0
dtype: float64
```

```
df[['City']].nunique()
```

```
City    52
dtype: int64
```

#filter rows where indiduals thought was suicidal

df[df['Have you ever had suicidal thoughts ?']== 'Yes']

Dhanraj Tungar

	id	Gender	Age	City	Profession	Academic Pressure	Work Pressure	CGPA	Study Satisfaction	Job Satisfaction	Sleep Duration	Dietary Habits	Degree	Have you ever had suicidal thoughts ?	Work/Study Hours	Fi
0	2	Male	33.0	Visakhapatnam	Student	5.0	0.0	8.97	2.0	0.0	5-6 hours	Healthy	B.Pharm	Yes	3.0	
3	30	Female	28.0	Varanasi	Student	3.0	0.0	5.59	2.0	0.0	7-8 hours	Moderate	BCA	Yes	4.0	
4	32	Female	25.0	Jaipur	Student	4.0	0.0	8.13	3.0	0.0	5-6 hours	Moderate	M.Tech	Yes	1.0	
8	59	Male	28.0	Nagpur	Student	3.0	0.0	9.79	1.0	0.0	7-8 hours	Moderate	B.Ed	Yes	12.0	
9	62	Male	31.0	Nashik	Student	2.0	0.0	8.38	3.0	0.0	Less than 5 hours	Moderate	LLB	Yes	2.0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
27894	140681	Male	23.0	Srinagar	Student	3.0	0.0	6.00	2.0	0.0	More than 8 hours	Healthy	MBBS	Yes	12.0	
27895	140684	Male	31.0	Lucknow	Student	2.0	0.0	7.27	5.0	0.0	7-8 hours	Moderate	B.Com	Yes	6.0	
27896	140685	Female	27.0	Surat	Student	5.0	0.0	5.75	5.0	0.0	5-6 hours	Unhealthy	Class 12	Yes	7.0	
27899	140690	Female	18.0	Ludhiana	Student	5.0	0.0	6.88	2.0	0.0	Less than 5 hours	Healthy	Class 12	Yes	10.0	
27900	140699	Male	27.0	Patna	Student	4.0	0.0	9.24	1.0	0.0	Less than 5 hours	Healthy	BCA	Yes	2.0	

17656 rows × 18 columns

```
#filter for high finacial stress
```

```
df[['Financial Stress']].max()
```

```
Financial Stress    5.0
dtype: float64
```

df[df['Financial Stress']>3]

Dhanraj Tungar

	id	Gender	Age	City	Profession	Academic Pressure	Work Pressure	CGPA	Study Satisfaction	Job Satisfaction	Sleep Duration	Dietary Habits	Degree	Have you ever had suicidal thoughts ?	Work/Study Hours	Finan St
3	30	Female	28.0	Varanasi	Student	3.0	0.0	5.59	2.0	0.0	7-8 hours	Moderate	BCA	Yes	4.0	
9	62	Male	31.0	Nashik	Student	2.0	0.0	8.38	3.0	0.0	Less than 5 hours	Moderate	LLB	Yes	2.0	
14	103	Female	19.0	Kalyan	Student	5.0	0.0	5.64	5.0	0.0	Less than 5 hours	Moderate	Class 12	Yes	4.0	
16	120	Male	25.0	Nashik	Student	5.0	0.0	6.51	2.0	0.0	Less than 5 hours	Unhealthy	M.Ed	Yes	2.0	
22	166	Female	25.0	Ahmedabad	Student	3.0	0.0	5.57	3.0	0.0	More than 8 hours	Unhealthy	MSc	Yes	10.0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
27883	140584	Female	22.0	Kanpur	Student	4.0	0.0	6.61	2.0	0.0	More than 8 hours	Unhealthy	M.Com	Yes	6.0	
27887	140624	Male	32.0	Rajkot	Student	4.0	0.0	9.19	1.0	0.0	5-6 hours	Healthy	MSc	No	7.0	
27891	140645	Female	28.0	Thane	Student	4.0	0.0	7.77	3.0	0.0	Less than 5 hours	Unhealthy	MSc	No	2.0	
27894	140681	Male	23.0	Srinagar	Student	3.0	0.0	6.00	2.0	0.0	More than 8 hours	Healthy	MBBS	Yes	12.0	
27899	140690	Female	18.0	Ludhiana	Student	5.0	0.0	6.88	2.0	0.0	Less than 5 hours	Healthy	Class 12	Yes	10.0	

12490 rows × 18 columns

df['Have you ever had suicidal thoughts ?'].map({'Yes':1 , 'No' :0 })

```
0        1
1        0
2        0
3        1
4        1
        ..
27896    1
27897    0
27898    0
27899    1
27900    1
Name: Have you ever had suicidal thoughts ?, Length: 27901, dtype: int64
```

```
df.iloc[0,1]
```

```
'Male'
```

```
df.iloc[0,0]
```

```
2
```

#find all rows where gender is male and dipression is one

df[(df['Gender'] == 'Male') & (df['Depression'] == 1)]

Dhanraj Tungar

:

	id	Gender	Age	City	Profession	Academic Pressure	Work Pressure	CGPA	Study Satisfaction	Job Satisfaction	Sleep Duration	Dietary Habits	Degree	Have you ever had suicidal thoughts ?	Work/Study Hours	F
0	2	Male	33.0	Visakhapatnam	Student	5.0	0.0	8.97	2.0	0.0	5-6 hours	Healthy	B.Pharm	Yes	3.0	
8	59	Male	28.0	Nagpur	Student	3.0	0.0	9.79	1.0	0.0	7-8 hours	Moderate	B.Ed	Yes	12.0	
9	62	Male	31.0	Nashik	Student	2.0	0.0	8.38	3.0	0.0	Less than 5 hours	Moderate	LLB	Yes	2.0	
10	83	Male	24.0	Nagpur	Student	3.0	0.0	6.10	3.0	0.0	5-6 hours	Moderate	Class 12	Yes	11.0	
12	94	Male	27.0	Kalyan	Student	5.0	0.0	7.04	1.0	0.0	Less than 5 hours	Moderate	M.Tech	No	10.0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
27881	140564	Male	27.0	Surat	Student	4.0	0.0	5.16	4.0	0.0	Less than 5 hours	Moderate	M.Pharm	Yes	6.0	
27884	140594	Male	18.0	Meerut	Student	5.0	0.0	6.25	3.0	0.0	7-8 hours	Unhealthy	Class 12	No	11.0	
27887	140624	Male	32.0	Rajkot	Student	4.0	0.0	9.19	1.0	0.0	5-6 hours	Healthy	MSc	No	7.0	
27888	140630	Male	19.0	Kolkata	Student	4.0	0.0	7.13	1.0	0.0	More than 8 hours	Moderate	Class 12	No	10.0	
27900	140699	Male	27.0	Patna	Student	4.0	0.0	9.24	1.0	0.0	Less than 5 hours	Healthy	BCA	Yes	2.0	

9115 rows × 18 columns

```
#select first 3 rows of the age and CGPA COLUMN
```

```
df.loc[:2,['Age','CGPA']]
```

	Age	CGPA
0	33.0	8.97
1	24.0	5.90
2	31.0	7.03

```
#change the sleep duration for the person with id = 1 change sleep duration 8 - 9
```

```
if(df['id'] == 2).any():
    df.loc[df['id'] == 2 , 'Sleep Duration'] = '8-9 hours'
    print(df.loc[df['id'] == 2 ])
else:
    print('No row with id == 2')
```

```
   id Gender   Age           City Profession  Academic Pressure  \
0   2   Male  33.0  Visakhapatnam    Student                5.0

   Work Pressure  CGPA  Study Satisfaction  Job Satisfaction Sleep Duration  \
0            0.0  8.97                 2.0               0.0      8-9 hours

  Dietary Habits   Degree Have you ever had suicidal thoughts ?  \
0        Healthy  B.Pharm                                  Yes

   Work/Study Hours  Financial Stress Family History of Mental Illness  \
0               3.0               1.0                               No

   Depression
0           1
```

df.loc[df['id'] == 2 , 'Sleep Duration'] = '8-9 hours'

print(df.loc[df['id'] == 2])

```
   id Gender   Age           City Profession  Academic Pressure  \
0   2   Male  33.0  Visakhapatnam    Student                5.0

   Work Pressure  CGPA  Study Satisfaction  Job Satisfaction Sleep Duration  \
0            0.0  8.97                 2.0               0.0      8-9 hours

  Dietary Habits   Degree Have you ever had suicidal thoughts ?  \
0        Healthy  B.Pharm                                  Yes

   Work/Study Hours  Financial Stress Family History of Mental Illness  \
0               3.0               1.0                                No

   Depression
0           1
```

```
#print the 1st and 5th row along with 2nd column
```

```
df.iloc[[0,5],2]
```

```
0    33.0
5    29.0
Name: Age, dtype: float64
```

```
df.loc[[0,5],'Degree']
```

```
0    B.Pharm
5        PhD
Name: Degree, dtype: object
```

df['City'].value_counts()

City	
Kalyan	1570
Srinagar	1372
Hyderabad	1340
Vasai-Virar	1290
Lucknow	1155
Thane	1139
Ludhiana	1111
Agra	1094
Surat	1078
Kolkata	1066
Jaipur	1036
Patna	1007
Visakhapatnam	969
Pune	968
Ahmedabad	951
Bhopal	934
Chennai	885
Meerut	825
Rajkot	816
Delhi	768
Bangalore	767
Ghaziabad	745
Mumbai	699
Vadodara	694
Varanasi	685
Nagpur	651
Indore	643
Kanpur	609
Nashik	547
Faridabad	461
Saanvi	2
Bhavna	2
City	2
Harsha	2
Kibara	1
Nandini	1
Nalini	1
Mihir	1
Nalyan	1
M.Com	1

```
ME                          1
Rashi                       1
Gaurav                      1
Reyansh                     1
Harsh                       1
Vaanya                      1
Mira                        1
Less than 5 Kalyan          1
3.0                         1
Less Delhi                  1
M.Tech                      1
Khaziabad                   1
Name: count, dtype: int64
```

df.loc[1:18:2,['Degree','Sleep Duration']]

	Degree	Sleep Duration
1	BSc	5-6 hours
3	BCA	7-8 hours
5	PhD	Less than 5 hours
7	Class 12	Less than 5 hours
9	LLB	Less than 5 hours
11	BE	Less than 5 hours
13	Class 12	Less than 5 hours
15	M.Tech	More than 8 hours
17	Class 12	5-6 hours

df.loc[1:18:2,'City':'CGPA']

	City	Profession	Academic Pressure	Work Pressure	CGPA
1	Bangalore	Student	2.0	0.0	5.90
3	Varanasi	Student	3.0	0.0	5.59
5	Pune	Student	2.0	0.0	5.70
7	Chennai	Student	2.0	0.0	8.04
9	Nashik	Student	2.0	0.0	8.38
11	Vadodara	Student	3.0	0.0	7.03
13	Rajkot	Student	2.0	0.0	8.52
15	Srinagar	Student	3.0	0.0	8.58
17	Ahmedabad	Student	5.0	0.0	7.25

- 2 nd Notebook (Class Test)

```
import pandas as pd
df = pd.read_csv('temp.csv')
df
```

	Unnamed: 0	Temperature	Wind Speed	Event
0	Monday	24.0	18.0	Sunny
1	Tuesday	28.0	19.0	Cloudy
2	Wednesday	30.0	NaN	Rainy
3	Thursday	NaN	20.0	Windy
4	Friday	18.0	10.0	NaN

```
#show first 3 row
df.head(3)
```

	Unnamed: 0	Temperature	Wind Speed	Event
0	Monday	24.0	18.0	Sunny
1	Tuesday	28.0	19.0	Cloudy
2	Wednesday	30.0	NaN	Rainy

```
#show last row
df.tail(1)
```

	Unnamed: 0	Temperature	Wind Speed	Event
4	Friday	18.0	10.0	NaN

```
#print second row value from temperature column
#iloc[2,df['Temperature']]
df.iloc[1,1]
```

```
28.0
```

#fill null values using fillna()

```
new_df = df.fillna({
    'Temperature' : 22,
    'Wind Speed' : 17,
    'Event' : 'No Event'
})
new_df
```

	Unnamed: 0	Temperature	Wind Speed	Event
0	Monday	24.0	18.0	Sunny
1	Tuesday	28.0	19.0	Cloudy
2	Wednesday	30.0	17.0	Rainy
3	Thursday	22.0	20.0	Windy
4	Friday	18.0	10.0	No Event

```
df.fillna(method='bfill')
```

```
C:\Users\DELL80\AppData\Local\Temp\ipykernel_8188\2831856154.py:1: FutureWarning: DataFrame.fillna with 'method' is deprecated and will raise in a future version. Use obj.ffill() or obj.bfill() instead.
  df.fillna(method='bfill')
```

	Unnamed: 0	Temperature	Wind Speed	Event
0	Monday	24.0	18.0	Sunny
1	Tuesday	28.0	19.0	Cloudy
2	Wednesday	30.0	20.0	Rainy
3	Thursday	18.0	20.0	Windy
4	Friday	18.0	10.0	NaN

Dhanraj Tungar

df.dropna()

	Unnamed: 0	Temperature	Wind Speed	Event
0	Monday	24.0	18.0	Sunny
1	Tuesday	28.0	19.0	Cloudy

df.replace(18,20)

	Unnamed: 0	Temperature	Wind Speed	Event
0	Monday	24.0	20.0	Sunny
1	Tuesday	28.0	19.0	Cloudy
2	Wednesday	30.0	NaN	Rainy
3	Thursday	NaN	20.0	Windy
4	Friday	20.0	10.0	NaN

df[['Wind Speed','Event']][df['Temperature'] == 28]

	Wind Speed	Event
1	19.0	Cloudy

- Faker and GroupBy

```
pip install faker
```

```
Collecting faker
  Downloading Faker-35.2.0-py3-none-any.whl.metadata (15 kB)
Requirement already satisfied: python-dateutil>=2.4 in c:\users\dell80\anaconda3\lib\site-packages (from faker) (2.9.0.post0)
Requirement already satisfied: typing-extensions in c:\users\dell80\anaconda3\lib\site-packages (from faker) (4.11.0)
Requirement already satisfied: six>=1.5 in c:\users\dell80\anaconda3\lib\site-packages (from python-dateutil>=2.4->faker) (1.16.0)
Downloading Faker-35.2.0-py3-none-any.whl (1.9 MB)
   ---------------------------------------- 0.0/1.9 MB ? eta -:--:--
   ---------------------------------------- 1.9/1.9 MB 9.6 MB/s eta 0:00:00
Installing collected packages: faker
Successfully installed faker-35.2.0
Note: you may need to restart the kernel to use updated packages.
```

```
import faker as f
```

```
import pandas as pd
import random as rm
```

```
fake =f.Faker()
data = {
   "First Name": ["Rohit","Virat","Hardik","Jasprit","Tilak","Axar","Rishab","Rahul"],
   "Last Name": ["Sharma","Kohli","Pandya","Bumrah","Varma","Patel","Pant","Lokesh"],
   "e-mail": [fake.email() for _ in range(8)],
   "Age": [rm.randint(18,40) for _ in range(8)]
}
df =pd.DataFrame(data)
print(df)
```

```
  First Name Last Name                        e-mail  Age
0      Rohit    Sharma       justinbryant@example.com   21
1      Virat     Kohli         tschneider@example.net   27
2     Hardik    Pandya      hamiltonshaun@example.org   32
3    Jasprit    Bumrah     youngcatherine@example.net   25
4      Tilak     Varma  doughertyjennifer@example.org   22
5       Axar     Patel             jack74@example.com   26
6     Rishab      Pant          shawkeith@example.net   24
7      Rahul    Lokesh      brianmckinney@example.com   32
```

```
print(fake.name())
print(fake.address())
print(fake.email())
```

```
print(fake.phone_number())
```

```
Dana Friedman
1082 Richard Plains Suite 055
Michaelfurt, ND 36063
hryan@example.net
001-385-577-4500x40052
```

```
data = []
for _ in range(10):
    record = {
        "Name":fake.name(),
        "Address":fake.address(),
        "Email":fake.email(),
        "Date of birth":fake.date_of_birth(),
        "Phone Number":fake.phone_number,
        "Company":fake.company()
    }
    data.append(record)

df = pd.DataFrame(data)
df
```

	Name	Address	Email	Date of birth	Phone Number	Company
0	Michael Baker	58511 Taylor Prairie Apt. 747\nWest David, PA ...	tylersmith@example.net	1927-01-11	<bound method Provider.phone_number of <faker....	Silva-Wright
1	John Crawford	21541 Chris Shores\nSouth Ronald, ND 79442	downssusan@example.org	1997-08-11	<bound method Provider.phone_number of <faker....	Lin Ltd
2	Mark Pham	5629 Lori Run\nWest Martha, MI 87264	kevinfuller@example.net	1922-02-19	<bound method Provider.phone_number of <faker....	Torres, Sherman and Mcdonald
3	Benjamin Warren	9674 Sheila Vista\nPort John, DE 19557	seanbartlett@example.org	2019-11-14	<bound method Provider.phone_number of <faker....	Wilson Group
4	Erica Williams	89926 Patricia Mall\nWest Danielle, IL 78042	john69@example.net	1910-05-18	<bound method Provider.phone_number of <faker....	Estes, Carroll and Frost
5	Roger Green	84639 Perez Plains Apt. 482\nJaclynfurt, DE 03914	mitchelljennifer@example.com	1959-03-04	<bound method Provider.phone_number of <faker....	White Group
6	Paula Hamilton	2203 Nicholas Valley Apt. 933\nWilliamstown, K...	brittanycobb@example.org	1970-01-12	<bound method Provider.phone_number of <faker....	Lewis, Taylor and Stevens
7	Kelly Atkins	9892 Bailey Curve\nNorth Melissaborough, UT 19113	morristiffany@example.com	1972-09-07	<bound method Provider.phone_number of <faker....	Pena, Allen and Hudson
8	Michael Bauer	028 Lucas Prairie\nNew Jeremyland, AR 72210	martinezlaura@example.com	1974-02-21	<bound method Provider.phone_number of <faker....	Schroeder PLC
9	Courtney Willis	439 Aimee Coves\nCooperborough, FM 47036	gilbertjennifer@example.com	1938-06-12	<bound method Provider.phone_number of <faker....	May, Matthews and Jackson

Dhanraj Tungar

fake1 = f.Faker('hi_IN')

fake1.name()

```
'शंकर खत्री'
```

```
fake1.name()
```

```
'सुभाष चवनजी'
```

```
df.columns
```

```
Index(['Name', 'Address', 'Email', 'Date of birth', 'Phone Number', 'Company'], dtype='object')
```

df.rename(columns={df.columns[0]:'Full Name'},inplace=True)

df

	Full Name	Address	Email	Date of birth	Phone Number	Company
0	Michael Baker	58511 Taylor Prairie Apt. 747\nWest David, PA ...	tylersmith@example.net	1927-01-11	<bound method Provider.phone_number of <faker....	Silva-Wright
1	John Crawford	21541 Chris Shores\nSouth Ronald, ND 79442	downssusan@example.org	1997-08-11	<bound method Provider.phone_number of <faker....	Lin Ltd
2	Mark Pham	5629 Lori Run\nWest Martha, MI 87264	kevinfuller@example.net	1922-02-19	<bound method Provider.phone_number of <faker....	Torres, Sherman and Mcdonald
3	Benjamin Warren	9674 Sheila Vista\nPort John, DE 19557	seanbartlett@example.org	2019-11-14	<bound method Provider.phone_number of <faker....	Wilson Group
4	Erica Williams	89926 Patricia Mall\nWest Danielle, IL 78042	john69@example.net	1910-05-18	<bound method Provider.phone_number of <faker....	Estes, Carroll and Frost
5	Roger Green	84639 Perez Plains Apt. 482\nJaclynfurt, DE 03914	mitchelljennifer@example.com	1959-03-04	<bound method Provider.phone_number of <faker....	White Group
6	Paula Hamilton	2203 Nicholas Valley Apt. 933\nWilliamstown, K...	brittanycobb@example.org	1970-01-12	<bound method Provider.phone_number of <faker....	Lewis, Taylor and Stevens
7	Kelly Atkins	9892 Bailey Curve\nNorth Melissaborough, UT 19113	morristiffany@example.com	1972-09-07	<bound method Provider.phone_number of <faker....	Pena, Allen and Hudson
8	Michael Bauer	028 Lucas Prairie\nNew Jeremyland, AR 72210	martinezlaura@example.com	1974-02-21	<bound method Provider.phone_number of <faker....	Schroeder PLC
9	Courtney Willis	439 Aimee Coves\nCooperborough, FM 47036	gilbertjennifer@example.com	1938-06-12	<bound method Provider.phone_number of <faker....	May, Matthews and Jackson

df.rename(columns={

"Phone Number":"Contact Number",

"Email":"Email Address"},

inplace=True

)

df

	Full Name	Address	Email Address	Date of birth	Contact Number	Company
0	Michael Baker	58511 Taylor Prairie Apt. 747\nWest David, PA ...	tylersmith@example.net	1927-01-11	<bound method Provider.phone_number of <faker....	Silva-Wright
1	John Crawford	21541 Chris Shores\nSouth Ronald, ND 79442	downssusan@example.org	1997-08-11	<bound method Provider.phone_number of <faker....	Lin Ltd
2	Mark Pham	5629 Lori Run\nWest Martha, MI 87264	kevinfuller@example.net	1922-02-19	<bound method Provider.phone_number of <faker....	Torres, Sherman and Mcdonald
3	Benjamin Warren	9674 Sheila Vista\nPort John, DE 19557	seanbartlett@example.org	2019-11-14	<bound method Provider.phone_number of <faker....	Wilson Group
4	Erica Williams	89926 Patricia Mall\nWest Danielle, IL 78042	john69@example.net	1910-05-18	<bound method Provider.phone_number of <faker....	Estes, Carroll and Frost
5	Roger Green	84639 Perez Plains Apt. 482\nJaclynfurt, DE 03914	mitchelljennifer@example.com	1959-03-04	<bound method Provider.phone_number of <faker....	White Group
6	Paula Hamilton	2203 Nicholas Valley Apt. 933\nWilliamstown, K...	brittanycobb@example.org	1970-01-12	<bound method Provider.phone_number of <faker....	Lewis, Taylor and Stevens
7	Kelly Atkins	9892 Bailey Curve\nNorth Melissaborough, UT 19113	morristiffany@example.com	1972-09-07	<bound method Provider.phone_number of <faker....	Pena, Allen and Hudson
8	Michael Bauer	028 Lucas Prairie\nNew Jeremyland, AR 72210	martinezlaura@example.com	1974-02-21	<bound method Provider.phone_number of <faker....	Schroeder PLC
9	Courtney Willis	439 Aimee Coves\nCooperborough, FM 47036	gilbertjennifer@example.com	1938-06-12	<bound method Provider.phone_number of <faker....	May, Matthews and Jackson

```
[x.upper() for x in df.columns]
```

```
['FULL NAME',
 'ADDRESS',
 'EMAIL ADDRESS',
 'DATE OF BIRTH',
 'CONTACT NUMBER',
 'COMPANY']
```

```
[x.lower() for x in df.columns]
```

```
['full name',
 'address',
 'email address',
 'date of birth',
 'contact number',
 'company']
```

Dhanraj Tungar

```
df
```

	Full Name	Address	Email Address	Date of birth	Contact Number	Company
0	Michael Baker	58511 Taylor Prairie Apt. 747\nWest David, PA ...	tylersmith@example.net	1927-01-11	<bound method Provider.phone_number of <faker....	Silva-Wright
1	John Crawford	21541 Chris Shores\nSouth Ronald, ND 79442	downssusan@example.org	1997-08-11	<bound method Provider.phone_number of <faker....	Lin Ltd
2	Mark Pham	5629 Lori Run\nWest Martha, MI 87264	kevinfuller@example.net	1922-02-19	<bound method Provider.phone_number of <faker....	Torres, Sherman and Mcdonald
3	Benjamin Warren	9674 Sheila Vista\nPort John, DE 19557	seanbartlett@example.org	2019-11-14	<bound method Provider.phone_number of <faker....	Wilson Group
4	Erica Williams	89926 Patricia Mall\nWest Danielle, IL 78042	john69@example.net	1910-05-18	<bound method Provider.phone_number of <faker....	Estes, Carroll and Frost
5	Roger Green	84639 Perez Plains Apt. 482\nJaclynfurt, DE 03914	mitchelljennifer@example.com	1959-03-04	<bound method Provider.phone_number of <faker....	White Group
6	Paula Hamilton	2203 Nicholas Valley Apt. 933\nWilliamstown, K...	brittanycobb@example.org	1970-01-12	<bound method Provider.phone_number of <faker....	Lewis, Taylor and Stevens
7	Kelly Atkins	9892 Bailey Curve\nNorth Melissaborough, UT 19113	morristiffany@example.com	1972-09-07	<bound method Provider.phone_number of <faker....	Pena, Allen and Hudson
8	Michael Bauer	028 Lucas Prairie\nNew Jeremyland, AR 72210	martinezlaura@example.com	1974-02-21	<bound method Provider.phone_number of <faker....	Schroeder PLC
9	Courtney Willis	439 Aimee Coves\nCooperborough, FM 47036	gilbertjennifer@example.com	1938-06-12	<bound method Provider.phone_number of <faker....	May, Matthews and Jackson

df.columns.str.replace(' ','_')

```
Index(['Full_Name', 'Address', 'Email_Address', 'Date_of_birth',
       'Contact_Number', 'Company'],
      dtype='object')
```

```
df
```

	Full Name	Address	Email Address	Date of birth	Contact Number	Company
0	Michael Baker	58511 Taylor Prairie Apt. 747\nWest David, PA ...	tylersmith@example.net	1927-01-11	<bound method Provider.phone_number of <faker....	Silva-Wright
1	John Crawford	21541 Chris Shores\nSouth Ronald, ND 79442	downssusan@example.org	1997-08-11	<bound method Provider.phone_number of <faker....	Lin Ltd
2	Mark Pham	5629 Lori Run\nWest Martha, MI 87264	kevinfuller@example.net	1922-02-19	<bound method Provider.phone_number of <faker....	Torres, Sherman and Mcdonald
3	Benjamin Warren	9674 Sheila Vista\nPort John, DE 19557	seanbartlett@example.org	2019-11-14	<bound method Provider.phone_number of <faker....	Wilson Group
4	Erica Williams	89926 Patricia Mall\nWest Danielle, IL 78042	john69@example.net	1910-05-18	<bound method Provider.phone_number of <faker....	Estes, Carroll and Frost
5	Roger Green	84639 Perez Plains Apt. 482\nJaclynfurt, DE 03914	mitchelljennifer@example.com	1959-03-04	<bound method Provider.phone_number of <faker....	White Group
6	Paula Hamilton	2203 Nicholas Valley Apt. 933\nWilliamstown, K...	brittanycobb@example.org	1970-01-12	<bound method Provider.phone_number of <faker....	Lewis, Taylor and Stevens
7	Kelly Atkins	9892 Bailey Curve\nNorth Melissaborough, UT 19113	morristiffany@example.com	1972-09-07	<bound method Provider.phone_number of <faker....	Pena, Allen and Hudson
8	Michael Bauer	028 Lucas Prairie\nNew Jeremyland, AR 72210	martinezlaura@example.com	1974-02-21	<bound method Provider.phone_number of <faker....	Schroeder PLC
9	Courtney Willis	439 Aimee Coves\nCooperborough, FM 47036	gilbertjennifer@example.com	1938-06-12	<bound method Provider.phone_number of <faker....	May, Matthews and Jackson

df.loc[2]=['Sujal','Satara','sujal@gmail.com','13thSep2004','7816524132','BFC']

Dhanraj Tungar

```
df
```

	Full Name	Address	Email Address	Date of birth	Contact Number	Company
0	Michael Baker	58511 Taylor Prairie Apt. 747\nWest David, PA ...	tylersmith@example.net	1927-01-11	<bound method Provider.phone_number of <faker....	Silva-Wright
1	John Crawford	21541 Chris Shores\nSouth Ronald, ND 79442	downssusan@example.org	1997-08-11	<bound method Provider.phone_number of <faker....	Lin Ltd
2	Sujal	Satara	sujal@gmail.com	13thSep2004	7816524132	BFC
3	Benjamin Warren	9674 Sheila Vista\nPort John, DE 19557	seanbartlett@example.org	2019-11-14	<bound method Provider.phone_number of <faker....	Wilson Group
4	Erica Williams	89926 Patricia Mall\nWest Danielle, IL 78042	john69@example.net	1910-05-18	<bound method Provider.phone_number of <faker....	Estes, Carroll and Frost
5	Roger Green	84639 Perez Plains Apt. 482\nJaclynfurt, DE 03914	mitchelljennifer@example.com	1959-03-04	<bound method Provider.phone_number of <faker....	White Group
6	Paula Hamilton	2203 Nicholas Valley Apt. 933\nWilliamstown, K...	brittanycobb@example.org	1970-01-12	<bound method Provider.phone_number of <faker....	Lewis, Taylor and Stevens
7	Kelly Atkins	9892 Bailey Curve\nNorth Melissaborough, UT 19113	morristiffany@example.com	1972-09-07	<bound method Provider.phone_number of <faker....	Pena, Allen and Hudson
8	Michael Bauer	028 Lucas Prairie\nNew Jeremyland, AR 72210	martinezlaura@example.com	1974-02-21	<bound method Provider.phone_number of <faker....	Schroeder PLC
9	Courtney Willis	439 Aimee Coves\nCooperborough, FM 47036	gilbertjennifer@example.com	1938-06-12	<bound method Provider.phone_number of <faker....	May, Matthews and Jackson

df.loc[2,['Address']]=['Airoli , New Mumbai']

```
df
```

	Full Name	Address	Email Address	Date of birth	Contact Number	Company
0	Michael Baker	58511 Taylor Prairie Apt. 747\nWest David, PA ...	tylersmith@example.net	1927-01-11	<bound method Provider.phone_number of <faker....	Silva-Wright
1	John Crawford	21541 Chris Shores\nSouth Ronald, ND 79442	downssusan@example.org	1997-08-11	<bound method Provider.phone_number of <faker....	Lin Ltd
2	Sujal	Airoli , New Mumbai	sujal@gmail.com	13thSep2004	7816524132	BFC
3	Benjamin Warren	9674 Sheila Vista\nPort John, DE 19557	seanbartlett@example.org	2019-11-14	<bound method Provider.phone_number of <faker....	Wilson Group
4	Erica Williams	89926 Patricia Mall\nWest Danielle, IL 78042	john69@example.net	1910-05-18	<bound method Provider.phone_number of <faker....	Estes, Carroll and Frost
5	Roger Green	84639 Perez Plains Apt. 482\nJaclynfurt, DE 03914	mitchelljennifer@example.com	1959-03-04	<bound method Provider.phone_number of <faker....	White Group
6	Paula Hamilton	2203 Nicholas Valley Apt. 933\nWilliamstown, K...	brittanycobb@example.org	1970-01-12	<bound method Provider.phone_number of <faker....	Lewis, Taylor and Stevens
7	Kelly Atkins	9892 Bailey Curve\nNorth Melissaborough, UT 19113	morristiffany@example.com	1972-09-07	<bound method Provider.phone_number of <faker....	Pena, Allen and Hudson
8	Michael Bauer	028 Lucas Prairie\nNew Jeremyland, AR 72210	martinezlaura@example.com	1974-02-21	<bound method Provider.phone_number of <faker....	Schroeder PLC
9	Courtney Willis	439 Aimee Coves\nCooperborough, FM 47036	gilbertjennifer@example.com	1938-06-12	<bound method Provider.phone_number of <faker....	May, Matthews and Jackson

df['Company'].apply(len)

```
0    12
1     7
2     3
3    12
4    24
5    11
6    25
7    22
8    13
9    25
Name: Company, dtype: int64
```

df.apply(len)

```
Full Name         10
Address           10
Email Address     10
Date of birth     10
Contact Number    10
Company           10
dtype: int64
```

```
len(df['Full Name'])
```

```
10
```

```
df.apply(len,axis='columns')
```

```
0    6
1    6
2    6
3    6
4    6
5    6
6    6
7    6
8    6
9    6
dtype: int64
```

```
def update_email(Email):
    return Email.upper()
df['Email Address'].apply(update_email)
```

```
0           TYLERSMITH@EXAMPLE.NET
1           DOWNSSUSAN@EXAMPLE.ORG
2                    SUJAL@GMAIL.COM
3         SEANBARTLETT@EXAMPLE.ORG
4                JOHN69@EXAMPLE.NET
5     MITCHELLJENNIFER@EXAMPLE.COM
6         BRITTANYCOBB@EXAMPLE.ORG
7       MORRISTIFFANY@EXAMPLE.COM
8       MARTINEZLAURA@EXAMPLE.COM
9     GILBERTJENNIFER@EXAMPLE.COM
Name: Email Address, dtype: object
```

```
df['Email Address'] = df['Email Address'].str.upper()
```

df

	Full Name	Address	Email Address	Date of birth	Contact Number	Company
0	Michael Baker	58511 Taylor Prairie Apt. 747\nWest David, PA ...	TYLERSMITH@EXAMPLE.NET	1927-01-11	<bound method Provider.phone_number of <faker....	Silva-Wright
1	John Crawford	21541 Chris Shores\nSouth Ronald, ND 79442	DOWNSSUSAN@EXAMPLE.ORG	1997-08-11	<bound method Provider.phone_number of <faker....	Lin Ltd
2	Sujal	Airoli , New Mumbai	SUJAL@GMAIL.COM	13thSep2004	7816524132	BFC
3	Benjamin Warren	9674 Sheila Vista\nPort John, DE 19557	SEANBARTLETT@EXAMPLE.ORG	2019-11-14	<bound method Provider.phone_number of <faker....	Wilson Group
4	Erica Williams	89926 Patricia Mall\nWest Danielle, IL 78042	JOHN69@EXAMPLE.NET	1910-05-18	<bound method Provider.phone_number of <faker....	Estes, Carroll and Frost
5	Roger Green	84639 Perez Plains Apt. 482\nJaclynfurt, DE 03914	MITCHELLJENNIFER@EXAMPLE.COM	1959-03-04	<bound method Provider.phone_number of <faker....	White Group
6	Paula Hamilton	2203 Nicholas Valley Apt. 933\nWilliamstown, K...	BRITTANYCOBB@EXAMPLE.ORG	1970-01-12	<bound method Provider.phone_number of <faker....	Lewis, Taylor and Stevens
7	Kelly Atkins	9892 Bailey Curve\nNorth Melissaborough, UT 19113	MORRISTIFFANY@EXAMPLE.COM	1972-09-07	<bound method Provider.phone_number of <faker....	Pena, Allen and Hudson
8	Michael Bauer	028 Lucas Prairie\nNew Jeremyland, AR 72210	MARTINEZLAURA@EXAMPLE.COM	1974-02-21	<bound method Provider.phone_number of <faker....	Schroeder PLC
9	Courtney Willis	439 Aimee Coves\nCooperborough, FM 47036	GILBERTJENNIFER@EXAMPLE.COM	1938-06-12	<bound method Provider.phone_number of <faker....	May, Matthews and Jackson

```
df['Email Address']=df['Email Address'].apply(lambda x:x.lower())
```

df

Dhanraj Tungar

	Full Name	Address	Email Address	Date of birth	Contact Number	Company
0	Michael Baker	58511 Taylor Prairie Apt. 747\nWest David, PA ...	tylersmith@example.net	1927-01-11	<bound method Provider.phone_number of <faker....	Silva-Wright
1	John Crawford	21541 Chris Shores\nSouth Ronald, ND 79442	downssusan@example.org	1997-08-11	<bound method Provider.phone_number of <faker....	Lin Ltd
2	Sujal	Airoli , New Mumbai	sujal@gmail.com	13thSep2004	7816524132	BFC
3	Benjamin Warren	9674 Sheila Vista\nPort John, DE 19557	seanbartlett@example.org	2019-11-14	<bound method Provider.phone_number of <faker....	Wilson Group
4	Erica Williams	89926 Patricia Mall\nWest Danielle, IL 78042	john69@example.net	1910-05-18	<bound method Provider.phone_number of <faker....	Estes, Carroll and Frost
5	Roger Green	84639 Perez Plains Apt. 482\nJaclynfurt, DE 03914	mitchelljennifer@example.com	1959-03-04	<bound method Provider.phone_number of <faker....	White Group
6	Paula Hamilton	2203 Nicholas Valley Apt. 933\nWilliamstown, K...	brittanycobb@example.org	1970-01-12	<bound method Provider.phone_number of <faker....	Lewis, Taylor and Stevens
7	Kelly Atkins	9892 Bailey Curve\nNorth Melissaborough, UT 19113	morristiffany@example.com	1972-09-07	<bound method Provider.phone_number of <faker....	Pena, Allen and Hudson
8	Michael Bauer	028 Lucas Prairie\nNew Jeremyland, AR 72210	martinezlaura@example.com	1974-02-21	<bound method Provider.phone_number of <faker....	Schroeder PLC
9	Courtney Willis	439 Aimee Coves\nCooperborough, FM 47036	gilbertjennifer@example.com	1938-06-12	<bound method Provider.phone_number of <faker....	May, Matthews and Jackson

df[['Full Name','Company']].applymap(len)

	Full Name	Company
0	13	12
1	13	7
2	5	3
3	15	12
4	14	24
5	11	11
6	14	25
7	12	22
8	13	13
9	15	25

df['Full Name']+' '+df['Email Address']

```
0          Michael Baker tylersmith@example.net
1          John Crawford downssusan@example.org
2                        Sujal sujal@gmail.com
3      Benjamin Warren seanbartlett@example.org
4             Erica Williams john69@example.net
5      Roger Green mitchelljennifer@example.com
6       Paula Hamilton brittanycobb@example.org
7        Kelly Atkins morristiffany@example.com
8       Michael Bauer martinezlaura@example.com
9   Courtney Willis gilbertjennifer@example.com
dtype: object
```

df['Details']=df['Full Name']+' '+df['Email Address']

df

	Full Name	Address	Email Address	Date of birth	Contact Number	Company	Details
0	Michael Baker	58511 Taylor Prairie Apt. 747\nWest David, PA ...	tylersmith@example.net	1927-01-11	<bound method Provider.phone_number of <faker....	Silva-Wright	Michael Baker tylersmith@example.net
1	John Crawford	21541 Chris Shores\nSouth Ronald, ND 79442	downssusan@example.org	1997-08-11	<bound method Provider.phone_number of <faker....	Lin Ltd	John Crawford downssusan@example.org
2	Sujal	Airoli , New Mumbai	sujal@gmail.com	13thSep2004	7816524132	BFC	Sujal sujal@gmail.com
3	Benjamin Warren	9674 Sheila Vista\nPort John, DE 19557	seanbartlett@example.org	2019-11-14	<bound method Provider.phone_number of <faker....	Wilson Group	Benjamin Warren seanbartlett@example.org
4	Erica Williams	89926 Patricia Mall\nWest Danielle, IL 78042	john69@example.net	1910-05-18	<bound method Provider.phone_number of <faker....	Estes, Carroll and Frost	Erica Williams john69@example.net
5	Roger Green	84639 Perez Plains Apt. 482\nJaclynfurt, DE 03914	mitchelljennifer@example.com	1959-03-04	<bound method Provider.phone_number of <faker....	White Group	Roger Green mitchelljennifer@example.com
6	Paula Hamilton	2203 Nicholas Valley Apt. 933\nWilliamstown, K...	brittanycobb@example.org	1970-01-12	<bound method Provider.phone_number of <faker....	Lewis, Taylor and Stevens	Paula Hamilton brittanycobb@example.org
7	Kelly Atkins	9892 Bailey Curve\nNorth Melissaborough, UT 19113	morristiffany@example.com	1972-09-07	<bound method Provider.phone_number of <faker....	Pena, Allen and Hudson	Kelly Atkins morristiffany@example.com
8	Michael Bauer	028 Lucas Prairie\nNew Jeremyland, AR 72210	martinezlaura@example.com	1974-02-21	<bound method Provider.phone_number of <faker....	Schroeder PLC	Michael Bauer martinezlaura@example.com
9	Courtney Willis	439 Aimee Coves\nCooperborough, FM 47036	gilbertjennifer@example.com	1938-06-12	<bound method Provider.phone_number of <faker....	May, Matthews and Jackson	Courtney Willis gilbertjennifer@example.com

df.drop(columns= 'Details')

Dhanraj Tungar

	Full Name	Address	Email Address	Date of birth	Contact Number	Company
0	Michael Baker	58511 Taylor Prairie Apt. 747\nWest David, PA ...	tylersmith@example.net	1927-01-11	<bound method Provider.phone_number of <faker....	Silva-Wright
1	John Crawford	21541 Chris Shores\nSouth Ronald, ND 79442	downssusan@example.org	1997-08-11	<bound method Provider.phone_number of <faker....	Lin Ltd
2	Sujal	Airoli , New Mumbai	sujal@gmail.com	13thSep2004	7816524132	BFC
3	Benjamin Warren	9674 Sheila Vista\nPort John, DE 19557	seanbartlett@example.org	2019-11-14	<bound method Provider.phone_number of <faker....	Wilson Group
4	Erica Williams	89926 Patricia Mall\nWest Danielle, IL 78042	john69@example.net	1910-05-18	<bound method Provider.phone_number of <faker....	Estes, Carroll and Frost
5	Roger Green	84639 Perez Plains Apt. 482\nJaclynfurt, DE 03914	mitchelljennifer@example.com	1959-03-04	<bound method Provider.phone_number of <faker....	White Group
6	Paula Hamilton	2203 Nicholas Valley Apt. 933\nWilliamstown, K...	brittanycobb@example.org	1970-01-12	<bound method Provider.phone_number of <faker....	Lewis, Taylor and Stevens
7	Kelly Atkins	9892 Bailey Curve\nNorth Melissaborough, UT 19113	morristiffany@example.com	1972-09-07	<bound method Provider.phone_number of <faker....	Pena, Allen and Hudson
8	Michael Bauer	028 Lucas Prairie\nNew Jeremyland, AR 72210	martinezlaura@example.com	1974-02-21	<bound method Provider.phone_number of <faker....	Schroeder PLC
9	Courtney Willis	439 Aimee Coves\nCooperborough, FM 47036	gilbertjennifer@example.com	1938-06-12	<bound method Provider.phone_number of <faker....	May, Matthews and Jackson

df['Details']=df['Full Name']+' '+df['Email Address']

df['Details'].str.split(' ',expand=True)

	0	1	2
0	Michael	Baker	tylersmith@example.net
1	John	Crawford	downssusan@example.org
2	Sujal	sujal@gmail.com	None
3	Benjamin	Warren	seanbartlett@example.org
4	Erica	Williams	john69@example.net
5	Roger	Green	mitchelljennifer@example.com
6	Paula	Hamilton	brittanycobb@example.org
7	Kelly	Atkins	morristiffany@example.com
8	Michael	Bauer	martinezlaura@example.com
9	Courtney	Willis	gilbertjennifer@example.com

```
fake = f.Faker('en_IN')

data = {

   "Name":[fake.name() for _ in range(10)],

   "Age":[rm.randint(18,35) for _ in range(10)],

   "City":[fake.address() for _ in range(10)],


"Hobby":["Cricket","Reading","Swimming","Cooking","Football","Cricket","Swimming","Rea
ding","Dancing","Swimming"],

   "Bank Name":["ICICI","SBI","MAHARASHTRA
BANK","HDFC","KOTAK","SBI","HDFC","ICICI","MAHARASHTRA BANK","PNB"],


"Gender":["Male","Male","Male","Female","Female","Female","Male","Female","Female","
Male"]


}

df=pd.DataFrame(data)

df
```

	Name	Age	City	Hobby	Bank Name	Gender
0	Qushi Kashyap	25	H.No. 89\nMani Nagar\nDewas-551215	Cricket	ICICI	Male
1	Gaurang Banerjee	27	H.No. 916\nRaghavan\nShivpuri 654261	Reading	SBI	Male
2	Chandani Yadav	31	H.No. 76\nRau Marg\nKota-300078	Swimming	MAHARASHTRA BANK	Male
3	Hritik Mander	22	37/522\nMajumdar Chowk\nKarnal-892533	Cooking	HDFC	Female
4	Kalpit Varghese	34	H.No. 321\nGola Street, Panchkula-208032	Football	KOTAK	Female
5	Pahal Shanker	18	21, Garde Marg, Kolhapur 167765	Cricket	SBI	Female
6	Owen Narain	32	H.No. 54, Uppal, Machilipatnam 692267	Swimming	HDFC	Male
7	Logan Gole	20	H.No. 02\nGupta Marg\nAkola 281252	Reading	ICICI	Female
8	Hemangini Dar	32	H.No. 90, Jayaraman\nRamgarh-679057	Dancing	MAHARASHTRA BANK	Female
9	Eta Buch	20	69/579, Bahl Circle, Agartala-889235	Swimming	PNB	Male

```
df.sort_values(by=['Hobby','City'])
```

Dhanraj Tungar

	Name	Age	City	Hobby	Bank Name	Gender
3	Hritik Mander	22	37/522\nMajumdar Chowk\nKarnal-892533	Cooking	HDFC	Female
5	Pahal Shanker	18	21, Garde Marg, Kolhapur 167765	Cricket	SBI	Female
0	Qushi Kashyap	25	H.No. 89\nMani Nagar\nDewas-551215	Cricket	ICICI	Male
8	Hemangini Dar	32	H.No. 90, Jayaraman\nRamgarh-679057	Dancing	MAHARASHTRA BANK	Female
4	Kalpit Varghese	34	H.No. 321\nGola Street, Panchkula-208032	Football	KOTAK	Female
7	Logan Gole	20	H.No. 02\nGupta Marg\nAkola 281252	Reading	ICICI	Female
1	Gaurang Banerjee	27	H.No. 916\nRaghavan\nShivpuri 654261	Reading	SBI	Male
9	Eta Buch	20	69/579, Bahl Circle, Agartala-889235	Swimming	PNB	Male
6	Owen Narain	32	H.No. 54, Uppal, Machilipatnam 692267	Swimming	HDFC	Male
2	Chandani Yadav	31	H.No. 76\nRau Marg\nKota-300078	Swimming	MAHARASHTRA BANK	Male

df.groupby('Gender')['Age'].mean()

```
Gender
Female    25.2
Male      27.0
Name: Age, dtype: float64
```

```
df.groupby('City')['Name'].count()
```

```
City
21, Garde Marg, Kolhapur 167765              1
37/522\nMajumdar Chowk\nKarnal-892533        1
69/579, Bahl Circle, Agartala-889235         1
H.No. 02\nGupta Marg\nAkola 281252           1
H.No. 321\nGola Street, Panchkula-208032     1
H.No. 54, Uppal, Machilipatnam 692267        1
H.No. 76\nRau Marg\nKota-300078              1
H.No. 89\nMani Nagar\nDewas-551215           1
H.No. 90, Jayaraman\nRamgarh-679057          1
H.No. 916\nRaghavan\nShivpuri 654261         1
Name: Name, dtype: int64
```

#test

data = {

"Name":[fake.name() for _ in range(10)],

"Age":[rm.randint(18,35) for _ in range(10)],

```
    "City":[fake.city() for _ in range(10)],

"Hobby":["Cricket","Reading","Swimming","Cooking","Football","Cricket","Swimming","Rea
ding","Dancing","Swimming"],
    "Bank Name":["ICICI","SBI","MAHARASHTRA
BANK","HDFC","KOTAK","SBI","HDFC","ICICI","MAHARASHTRA BANK","PNB"],

"Gender":["Male","Male","Male","Female","Female","Female","Male","Female","Female","
Male"]


}
#test
data = {
    "Name":[fake.name() for _ in range(10)],
    "Age":[rm.randint(18,35) for _ in range(10)],
    "City":[fake.city() for _ in range(10)],

"Hobby":["Cricket","Reading","Swimming","Cooking","Football","Cricket","Swimming","Rea
ding","Dancing","Swimming"],
    "Bank Name":["ICICI","SBI","MAHARASHTRA
BANK","HDFC","KOTAK","SBI","HDFC","ICICI","MAHARASHTRA BANK","PNB"],

"Gender":["Male","Male","Male","Female","Female","Female","Male","Female","Female","
Male"]


}
df1
```

Dhanraj Tungar

	Name	Age	City	Hobby	Bank Name	Gender
0	Warda Ganesh	18	Thane	Cricket	ICICI	Male
1	Baghyawati Anne	30	Orai	Reading	SBI	Male
2	Neel Dara	32	New Delhi	Swimming	MAHARASHTRA BANK	Male
3	Alexander Kannan	29	Khandwa	Cooking	HDFC	Female
4	Aarnav Chatterjee	26	Dehradun	Football	KOTAK	Female
5	Ishita Koshy	34	Bharatpur	Cricket	SBI	Female
6	Aarna Jha	25	Ichalkaranji	Swimming	HDFC	Male
7	Dayita Hegde	20	Tiruppur	Reading	ICICI	Female
8	Darpan Pau	22	Yamunanagar	Dancing	MAHARASHTRA BANK	Female
9	Gautami Sampath	25	Pudukkottai	Swimming	PNB	Male

df.groupby('Bank Name')['Age'].mean()

```
Bank Name
HDFC                27.0
ICICI               19.0
KOTAK               26.0
MAHARASHTRA BANK    27.0
PNB                 25.0
SBI                 32.0
Name: Age, dtype: float64
```

df.groupby('Gender')['Age'].sum()

```
Gender
Female    131
Male      130
Name: Age, dtype: int64
```

```
df.groupby('Gender').agg(
   {
   'Age': ['mean','sum'],
   'Hobby': 'count',
   'Bank Name': 'nunique'
   }
)
```

	Age	Hobby	Bank Name
	mean	sum	count	nunique
Gender				
Female	26.2	131	5	5
Male	26.0	130	5	5

filt = df1['City'] != 'Mumbai'

df.loc[filt]['Hobby'].value_counts()

```
Hobby
Swimming    3
Cricket     2
Reading     2
Cooking     1
Football    1
Dancing     1
Name: count, dtype: int64
```

```
filt = df1['City'] == 'Thane'
df.loc[filt]['Hobby'].value_counts()
```

```
Hobby
Cricket    1
Name: count, dtype: int64
```

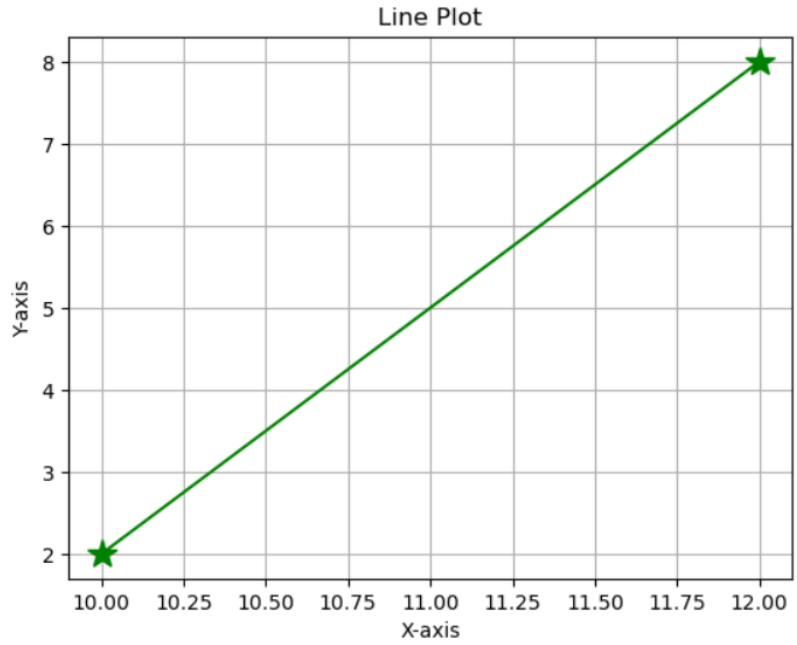
```
filt = df['Age'] < 24
df.loc[filt]['City'].str.contains('Thane').sum()
```

```
1
```

- # Graph

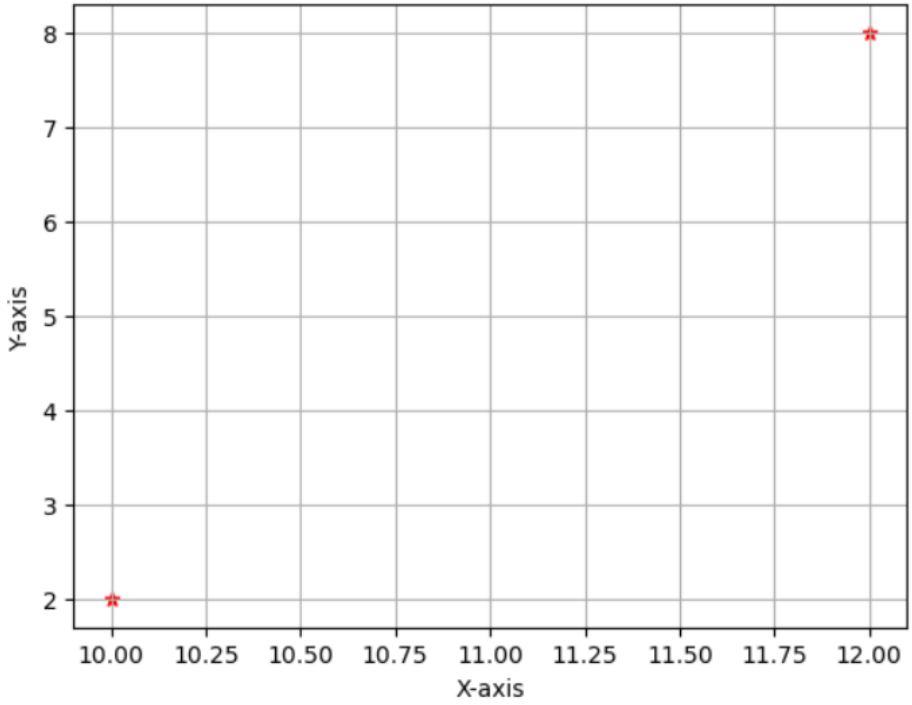
```
import pandas as pd
import matplotlib.pyplot as plt
```

```
x=[10,12]
y=[2,8]
plt.plot(x,y,marker="*",color='green',linestyle='-',markersize=15)
plt.title('Line Plot')
plt.xlabel('X-axis')
plt.ylabel('Y-axis')
plt.grid()
plt.show()
```

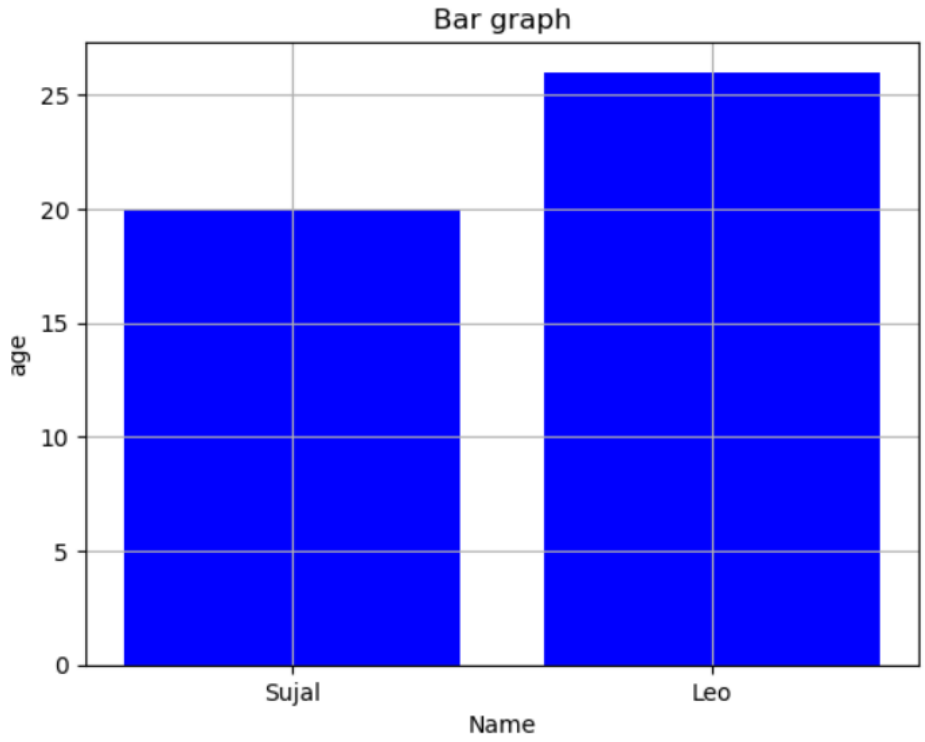



```
x=[10,12]
y=[2,8]
plt.scatter(x,y,marker="*",color='red')
plt.xlabel('X-axis')
plt.ylabel('Y-axis')
plt.grid()
```

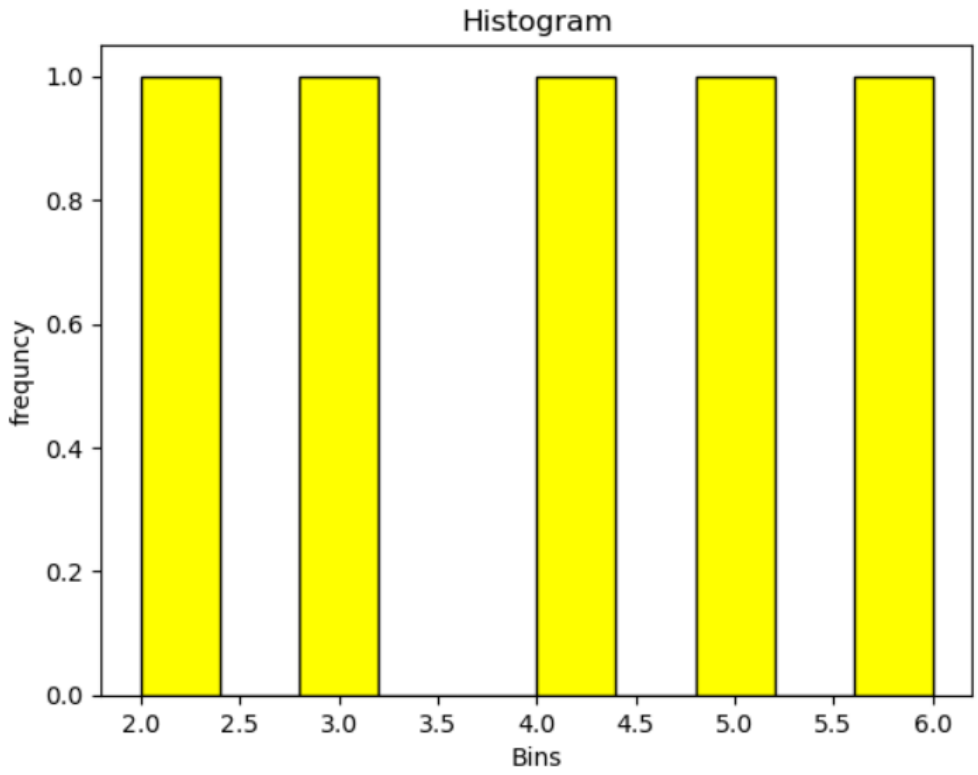
```
plt.show()
```




```
Name=['Sujal','Leo']
age=[20,26]
plt.bar(Name,age,color='blue')
plt.title('Bar graph')
plt.xlabel('Name')
plt.ylabel('age')
plt.grid()
plt.show()
```




```
data = [2,3,4,5,6]

plt.hist(data,bins= 10,color='Yellow',edgecolor='black')

plt.title('Histogram')

plt.xlabel('Bins')

plt.ylabel('frequncy')

plt.show()
```




```
sizes = [20,34,32,45,56]
labels = ['a','b','c','d','e']
plt.pie(sizes,labels=labels, autopct='%1.1f%%',startangle=0)
plt.title("Pie chart")
plt.show()
```