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


#dataset source = "https://www.kaggle.com/uciml/student-alcohol-consumption"
df = pd.read_csv('student-mat.csv')


df.head()
```

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob
0	GP	F	18	U	GT3	A	4	4	at_home	teacher
1	GP	F	17	U	GT3	T	1	1	at_home	other
2	GP	F	15	U	LE3	T	1	1	at_home	other
3	GP	F	15	U	GT3	T	4	2	health	services
4	GP	F	16	U	GT3	T	3	3	other	other

5 rows × 33 columns

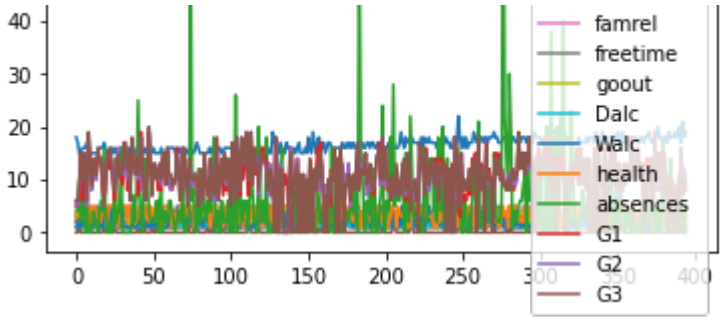
```
df.fillna(method='ffill', inplace=True)


df.describe()
```

	age	Medu	Fedu	traveltime	studytime	failure
count	395.000000	395.000000	395.000000	395.000000	395.000000	395.00000
mean	16.696203	2.749367	2.521519	1.448101	2.035443	0.33417
std	1.276043	1.094735	1.088201	0.697505	0.839240	0.74365
min	15.000000	0.000000	0.000000	1.000000	1.000000	0.00000
25%	16.000000	2.000000	2.000000	1.000000	1.000000	0.00000
50%	17.000000	3.000000	2.000000	1.000000	2.000000	0.00000
75%	18.000000	4.000000	3.000000	2.000000	2.000000	0.00000
max	22.000000	4.000000	4.000000	4.000000	4.000000	3.00000

```
df.plot()
```

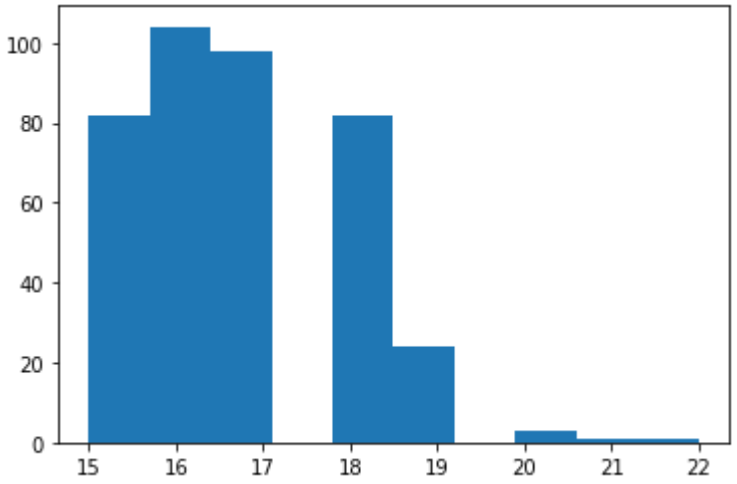
```
<AxesSubplot:>
```




```
x = df['age']
y = df['studytime']
```

```
plt.hist(x)
```

```
(array([ 82., 104.,  98.,   0.,  82.,  24.,   0.,   3.,   1.,   1.]),
 array([15. , 15.7, 16.4, 17.1, 17.8, 18.5, 19.2, 19.9, 20.6, 21.3, 22. ]),
 <BarContainer object of 10 artists>)
```



```
df.shape
```

```
(395, 33)
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 395 entries, 0 to 394
Data columns (total 33 columns):
 #   Column      Non-Null Count  Dtype 
---  ------      --------------  ----- 
 0   school      395 non-null    object
 1   sex         395 non-null    object
 2   age         395 non-null    int64 
 3   address     395 non-null    object
 4   famsize     395 non-null    object
 5   Pstatus     395 non-null    object
 6   Medu        395 non-null    int64 
 7   Fedu        395 non-null    int64 
 8   Mjob        395 non-null    object
```

```
 9   Fjob        395 non-null    object
 10  reason      395 non-null    object
 11  guardian    395 non-null    object
 12  traveltime  395 non-null    int64
 13  studytime   395 non-null    int64
 14  failures    395 non-null    int64
 15  schoolsup   395 non-null    object
 16  famsup      395 non-null    object
 17  paid        395 non-null    object
 18  activities  395 non-null    object
 19  nursery     395 non-null    object
 20  higher      395 non-null    object
 21  internet    395 non-null    object
 22  romantic    395 non-null    object
 23  famrel      395 non-null    int64
 24  freetime    395 non-null    int64
 25  goout       395 non-null    int64
 26  Dalc        395 non-null    int64
 27  Walc        395 non-null    int64
 28  health      395 non-null    int64
 29  absences    395 non-null    int64
 30  G1          395 non-null    int64
 31  G2          395 non-null    int64
 32  G3          395 non-null    int64
dtypes: int64(16), object(17)
memory usage: 102.0+ KB
```

```
x = df.traveltime
```

```
y = df.studytime
```

```
plt.hist(x)
```

```
(array([257.,   0.,   0., 107.,   0.,   0.,  23.,   0.,   0.,   8.]),
 array([1. , 1.3, 1.6, 1.9, 2.2, 2.5, 2.8, 3.1, 3.4, 3.7, 4. ]),
 <BarContainer object of 10 artists>)
```