Ex3 - Apply Functions

Step 1. Import the necessary libraries

In [1]:
```
import pandas as pd
```

Step 2. Read from student-mat.csv and assign it to a variable called df.

In [2]:
```
csv_url = 'student-mat.csv'
df = pd.read_csv(csv_url)
df.head()
```

Out[2]:

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	...	famrel
0	GP	F	18	U	GT3	A	4	4	at_home	teacher	...	4
1	GP	F	17	U	GT3	T	1	1	at_home	other	...	5
2	GP	F	15	U	LE3	T	1	1	at_home	other	...	4
3	GP	F	15	U	GT3	T	4	2	health	services	...	3
4	GP	F	16	U	GT3	T	3	3	other	other	...	4

5 rows × 33 columns

Step 3. Create a function called majority that returns a boolean value to a new column called legal_drinker (Consider majority as older than 17 years old)

In [3]:
```
def majority(x):
    if x > 17:
        return True
    else:
        return False
```

In [4]:
```
df['legal_drinker'] = df['age'].apply(majority)
df.head()
```

Out[4]:

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	...	freetim
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 × 34 columns

Ex 4 - Plotting

Step 1. Import the necessary libraries

In [5]:
```
import pandas as pd
import matplotlib.pyplot as plt
```

Step 2. Create the DataFrame it should look like below.

In [6]:
```
raw_data = {'first_name': ['Jason', 'Molly', 'Tina', 'Jake', 'Amy','John'],
            'last_name': ['Miller', 'Jacobson', 'Ali', 'Milner', 'Cooze','Co
            'female': [0, 1, 1, 0, 1,0],
            'age': [42, 52, 36, 42, 42,36],
            'preTestScore': [4, 24, 31, 2, 3, 50],
            'postTestScore': [25, 94, 57, 62, 70, 80]}

df = pd.DataFrame(raw_data)

df
```

Out[6]:

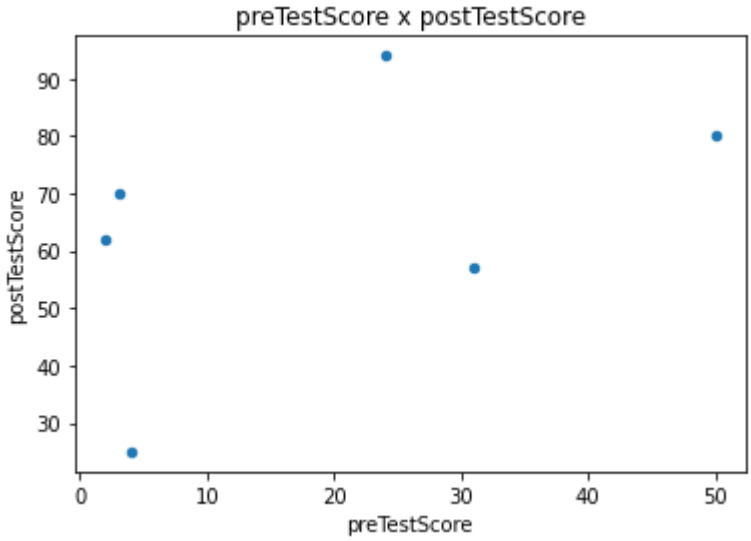
	first_name	last_name	female	age	preTestScore	postTestScore
0	Jason	Miller	0	42	4	25
1	Molly	Jacobson	1	52	24	94
2	Tina	Ali	1	36	31	57
3	Jake	Milner	0	42	2	62
4	Amy	Cooze	1	42	3	70
5	John	Cooper	0	36	50	80

Step 3. Create a Scatterplot with "preTestScore" as the x-axis and "postTestScore" as the y-axis.

In [7]:
```
df.plot(kind='scatter',x='preTestScore',y='postTestScore')

#set labels and titles
plt.title("preTestScore x postTestScore")
plt.xlabel('preTestScore')
plt.ylabel('postTestScore')

plt.show()
```



Step 4. Create a Histogram using column "Age" to find out the age distribution of participants in dataset.

In [8]:
```
df['age'].plot(kind='hist')

#set labels and titles
plt.title("Age Distribution")
plt.xlabel('Age')
plt.ylabel('Number of people')

plt.show()
```