

In this page, we will discover and analyze our dataset

```
In [2]: import pandas as pd
```

```
In [3]: my_df= pd.read_csv("who_suicide_statistics.csv")
```

```
In [4]: print(my_df.head(5))
```

	country	year	sex	age	suicides_no	population
0	Albania	1985	female	15-24 years	NaN	277900.0
1	Albania	1985	female	25-34 years	NaN	246800.0
2	Albania	1985	female	35-54 years	NaN	267500.0
3	Albania	1985	female	5-14 years	NaN	298300.0
4	Albania	1985	female	55-74 years	NaN	138700.0

```
In [15]: print(my_df.tail(5))
```

	country	year	sex	age	suicides_no	population
43771	Zimbabwe	1990	male	25-34 years	150.0	NaN
43772	Zimbabwe	1990	male	35-54 years	132.0	NaN
43773	Zimbabwe	1990	male	5-14 years	6.0	NaN
43774	Zimbabwe	1990	male	55-74 years	74.0	NaN
43775	Zimbabwe	1990	male	75+ years	13.0	NaN

```
In [5]: my_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 43776 entries, 0 to 43775
Data columns (total 6 columns):
country      43776 non-null object
year         43776 non-null int64
sex          43776 non-null object
age          43776 non-null object
suicides_no  41520 non-null float64
population   38316 non-null float64
dtypes: float64(2), int64(1), object(3)
memory usage: 2.0+ MB
```

```
In [6]: my_df.columns
```

```
Out[6]: Index(['country', 'year', 'sex', 'age', 'suicides_no', 'population'], dtype='object')
```

```
In [7]: my_df.index
```

```
Out[7]: RangeIndex(start=0, stop=43776, step=1)
```

In [8]: `my_df.describe()`

Out[8]:

	year	suicides_no	population
count	43776.000000	41520.000000	3.831600e+04
mean	1998.502467	193.315390	1.664091e+06
std	10.338711	800.589926	3.647231e+06
min	1979.000000	0.000000	2.590000e+02
25%	1990.000000	1.000000	8.511275e+04
50%	1999.000000	14.000000	3.806550e+05
75%	2007.000000	91.000000	1.305698e+06
max	2016.000000	22338.000000	4.380521e+07

In [9]: `import matplotlib.pyplot as plt`
`%matplotlib inline`

In [10]: `my_dfive = my_df[['age', 'population']].head()`

In [11]: `my_dfive`

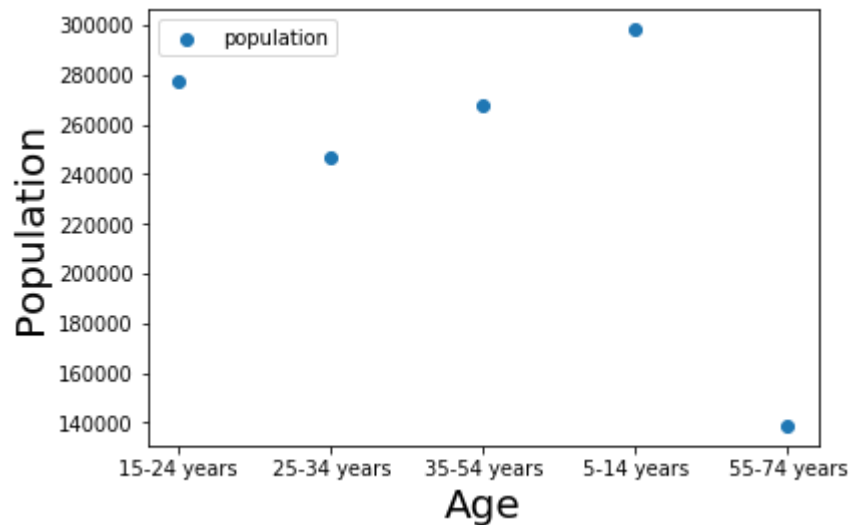
Out[11]:

	age	population
0	15-24 years	277900.0
1	25-34 years	246800.0
2	35-54 years	267500.0
3	5-14 years	298300.0
4	55-74 years	138700.0

```
In [12]: x= my_dfive.age
y= my_dfive.population

plt.xlabel("Age", fontsize=20)
plt.ylabel("Population", fontsize=20)
plt.title=('Albania Suicide Visualization')

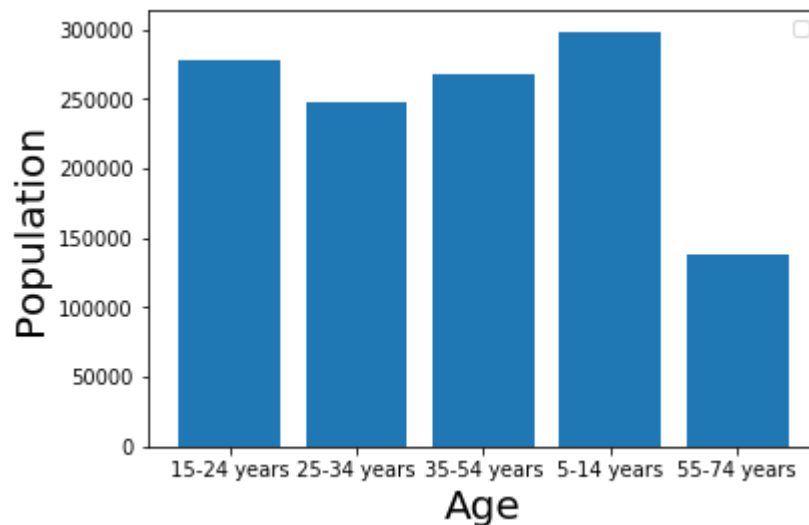
plt.scatter(x, y)
plt.legend()
plt.show()
```



```
In [13]: plt.xlabel("Age", fontsize=20)
plt.ylabel("Population", fontsize=20)
plt.title=('Albania Suicide Visualization')

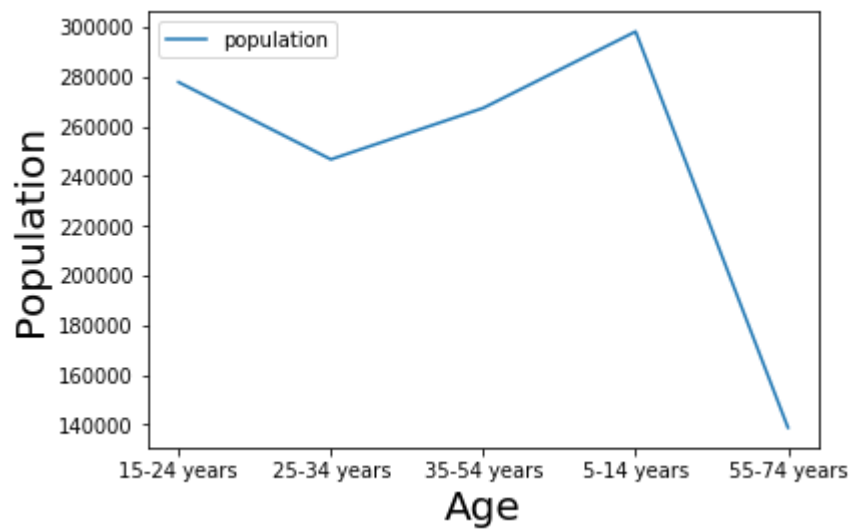
plt.bar(x, y)
plt.legend()
plt.show()
```

No handles with labels found to put in legend.



```
In [14]: plt.xlabel("Age", fontsize=20)
plt.ylabel("Population", fontsize=20)
plt.title('Albania Suicide Visualization')

plt.plot(x, y)
plt.legend()
plt.show()
```



```
In [37]: #filtering Data
my_cdf=my_df[pd.notna(my_df.population) & pd.notna(my_df.suicides_no)]
```

```
In [38]: #Grouping Data and Sum  
my_cdf.groupby(['country','sex'])['suicides_no'].sum()
```

```

Out[38]: country      sex      value
Albania      female    693.0
             male    1277.0
Antigua and Barbuda  female     1.0
             male    10.0
Argentina    female  21391.0
             male  72471.0
Armenia      female   655.0
             male  1767.0
Aruba        female   24.0
             male   96.0
Australia    female  17879.0
             male  62400.0
Austria      female  16150.0
             male  44029.0
Azerbaijan   female   991.0
             male  2375.0
Bahamas      female   16.0
             male   91.0
Bahrain      female   80.0
             male  383.0
Barbados     female   37.0
             male  168.0
Belarus      female  13055.0
             male  61919.0
Belgium      female  22588.0
             male  53360.0
Belize       female   51.0
             male  301.0
Bermuda      female    1.0
             male    5.0
...
Sweden       female  11186.0
             male  26614.0
Switzerland  female   7739.0
             male  18478.0
TFYR Macedonia  female  1041.0
             male   2436.0
Thailand      female  37173.0
             male  92726.0
Trinidad and Tobago  female   930.0
             male  3517.0
Turkey       female  2569.0
             male  7562.0
Turkmenistan  female  2414.0
             male  7230.0
Ukraine      female  71609.0
             male 293561.0
United Arab Emirates  female    98.0
             male   524.0
United Kingdom  female  43379.0
             male 122935.0
United States of America  female 253609.0
             male 947792.0
Uruguay      female   3088.0
             male  11559.0
Uzbekistan    female  11997.0

```

	male	32066.0
Venezuela (Bolivarian Republic of)	female	5300.0
	male	24204.0
Virgin Islands (USA)	female	19.0
	male	94.0

Name: suicides_no, Length: 236, dtype: float64

```
In [53]: #count of Country  
my_cdf['country'].nunique()
```

Out[53]: 118