

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

```
In [2]: df_suicide = pd.read_csv("suicide-rates-1985-to-2016-dataset.csv")
df_suicide.head()
```

Out[2]:

	country	year	sex	age	suicides_no	population	suicides/100k pop	country-year	HDI for year	gdp_for_year (\$)	gdp_per_capita (\$)	generation
0	Albania	1987	male	15-24 years	21	312900	6.71	Albania1987	NaN	2,156,624,900	796	Generation X
1	Albania	1987	male	35-54 years	16	308000	5.19	Albania1987	NaN	2,156,624,900	796	Silent
2	Albania	1987	female	15-24 years	14	289700	4.83	Albania1987	NaN	2,156,624,900	796	Generation X
3	Albania	1987	male	75+ years	1	21800	4.59	Albania1987	NaN	2,156,624,900	796	G.I. Generation
4	Albania	1987	male	25-34 years	9	274300	3.28	Albania1987	NaN	2,156,624,900	796	Boomers

```
In [3]: df_suicide = df_suicide.rename(columns = {'suicides/100k pop': 'suicides_per_100k_pop', \
                                                'gdp_for_year ($)': 'gdp_for_year', \
                                                'gdp_per_capita ($)': 'gdp_per_capita', \
                                                'HDI for year': 'hdi_for_year', \
                                                'country-year': 'country_year'}, inplace = False)
df_suicide.loc[df_suicide['country'] == 'United States'].head(1)
```

Out[3]:

	country	year	sex	age	suicides_no	population	suicides_per_100k_pop	country_year	hdi_for_year	gdp_for_year	gdp_per_capita	generation
26848	United States	1985	male	75+ years	2177	4064000	53.57	United States1985	0.841	4,346,734,000,000	19693	G.I. Generation

```
In [4]: # calculate the average suicide per 100k population by grouping country and finding the average
average_suicides_per_100k_pop = df_suicide[['country', 'suicides_per_100k_pop']].groupby(['country']).mean()
average_suicides_per_100k_pop.reset_index(inplace=True)
average_suicides_per_100k_pop = average_suicides_per_100k_pop.rename(columns = {'suicides_per_100k_pop': 'average_suicides_per_100k_pop'}, inplace = False)
df_suicide = df_suicide.merge(average_suicides_per_100k_pop[['country', 'average_suicides_per_100k_pop']])
df_suicide.head(3)
```

Out[4]:

	country	year	sex	age	suicides_no	population	suicides_per_100k_pop	country_year	hdi_for_year	gdp_for_year	gdp_per_capita	generation	average_suicides_per_100k_pop
0	Albania	1987	male	15-24 years	21	312900	6.71	Albania1987	NaN	2,156,624,900	796	Generation X	3.502879
1	Albania	1987	male	35-54 years	16	308000	5.19	Albania1987	NaN	2,156,624,900	796	Silent	3.502879
2	Albania	1987	female	15-24 years	14	289700	4.83	Albania1987	NaN	2,156,624,900	796	Generation X	3.502879

```
In [5]: df_suicide.to_csv("suicide-rates-1985-to-2016-dataset-cleaned.csv", index=False)
```