Colab: https://colab.research.google.com/drive/1fkgXd0C2Nh_hPybBxn1oV9MU4TPEVSmN?usp=sharing

```
import pandas as pd
```

```
import numpy as np
```

https://drive.google.com/file/d/1E3bwvYGf1ig32RmcYiWc0IXPN-mD_bI_/view?usp=sharing

```
!gdown 1E3bwvYGf1ig32RmcYiWc0IXPN-mD_bI_
```

```
Downloading...
From: https://drive.google.com/uc?id=1E3bwvYGf1ig32RmcYiWc0IXPN-mD_bI_
To: /content/gapminder.csv
100% 83.8k/83.8k [00:00<00:00, 63.8MB/s]
```

```
df = pd.read_csv("gapminder.csv")
df
```

	country	year	population	continent	life_exp	gdp_cap
			25333	Asia	28.801	779.445314
1	Afghanistan	1957	9240934	Asia	30.332	820.853030
2	Afghanistan	1962	10267083	Asia	31.997	853.100710
3	Afghanistan	1967	11537966	Asia	34.020	836.197138
4	Afghanistan	1972	13079460	Asia	36.088	739.981106
...	...	...	...	...	...	...
1699	Zimbabwe	1987	9216418	Africa	62.351	706.157306
1700	Zimbabwe	1992	10704340	Africa	60.377	693.420786
1701	Zimbabwe	1997	11404948	Africa	46.809	792.449960
1702	Zimbabwe	2002	11926563	Africa	39.989	672.038623
1703	Zimbabwe	2007	12311143	Africa	43.487	469.709298

1704 rows × 6 columns

```
type(df)
```

```
pandas.core.frame.DataFrame
```

```
df["population"]
```

```
0        8425333
1        9240934
2       10267083
3       11537966
4       13079460
          ...
1699     9216418
1700    10704340
1701    11404948
1702    11926563
1703    12311143
Name: population, Length: 1704, dtype: int64
```

```
type(df["population"])
```

```
pandas.core.series.Series
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1704 entries, 0 to 1703
Data columns (total 6 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   country     1704 non-null   object
 1   year        1704 non-null   int64
 2   population  1704 non-null   int64
 3   continent   1704 non-null   object
 4   life_exp    1704 non-null   float64
 5   gdp_cap     1704 non-null   float64
```

```
dtypes: float64(2), int64(2), object(2)
memory usage: 80.0+ KB
```

```
df.head(7)
```

	country	year	population	continent	life_exp	gdp_cap
0	Afghanistan	1952	8425333	Asia	28.801	779.445314
1	Afghanistan	1957	9240934	Asia	30.332	820.853030
2	Afghanistan	1962	10267083	Asia	31.997	853.100710
3	Afghanistan	1967	11537966	Asia	34.020	836.197138
4	Afghanistan	1972	13079460	Asia	36.088	739.981106
5	Afghanistan	1977	14880372	Asia	38.438	786.113360
6	Afghanistan	1982	12881816	Asia	39.854	978.011439

```
df.tail()
```

	country	year	population	continent	life_exp	gdp_cap
			418	Africa	62.351	706.157306
1700	Zimbabwe	1992	10704340	Africa	60.377	693.420786
1701	Zimbabwe	1997	11404948	Africa	46.809	792.449960
1702	Zimbabwe	2002	11926563	Africa	39.989	672.038623
1703	Zimbabwe	2007	12311143	Africa	43.487	469.709298

Saving...

```
df.shape
```

```
(1704, 6)
```

```
df.head(3)
```

	country	year	population	continent	life_exp	gdp_cap
0	Afghanistan	1952	8425333	Asia	28.801	779.445314
1	Afghanistan	1957	9240934	Asia	30.332	820.853030
2	Afghanistan	1962	10267083	Asia	31.997	853.100710

```
# A-1: Row oriented
# A-2: Column Oriented
```

```
pd.DataFrame([['Afghanistan',1952, 8425333, 'Asia', 28.801, 779.445314 ],
              ['Afghanistan',1957, 9240934, 'Asia', 30.332, 820.853030 ],
              ['Afghanistan',1962, 102267083, 'Asia', 31.997, 853.100710]],
             columns=['country','year','population','continent','life_exp','gdp_cap'])
```

	country	year	population	continent	life_exp	gdp_cap
0	Afghanistan	1952	8425333	Asia	28.801	779.445314
1	Afghanistan	1957	9240934	Asia	30.332	820.853030
2	Afghanistan	1962	102267083	Asia	31.997	853.100710

```
pd.DataFrame([['Afghanistan',1952, 8425333, 'Asia', 28.801, 779.445314]],
             columns=['country','year','population','continent','life_exp','gdp_cap'])
```

	country	year	population	continent	life_exp	gdp_cap
0	Afghanistan	1952	8425333	Asia	28.801	779.445314

```
pd.DataFrame({'country':['Afghanistan', 'Afghanistan'], 'year':[1952,1957],
              'population':[842533, 9240934], 'continent':['Asia', 'Asia'],
              'life_exp':[28.801, 30.332], 'gdp_cap':[779.445314, 820.853030]})
```

```
# Basic Ops on columns
```

```
df.columns
```

```
Index(['country', 'year', 'population', 'continent', 'life_exp', 'gdp_cap'], dtype='object')
```

```
df.keys()
```

```
Index(['country', 'year', 'population', 'continent', 'life_exp', 'gdp_cap'], dtype='object')
```

```
df[["country"]] # now this is a dataframe
```

	country
0	Afghanistan
1	Afghanistan
2	Afghanistan
3	Afghanistan

Saving...

1699	Zimbabwe
1700	Zimbabwe
1701	Zimbabwe
1702	Zimbabwe
1703	Zimbabwe

1704 rows × 1 columns

```
df[["country", "population"]]
```

	country	population
0	Afghanistan	8425333
1	Afghanistan	9240934
2	Afghanistan	10267083
3	Afghanistan	11537966
4	Afghanistan	13079460
...	...	...
1699	Zimbabwe	9216418
1700	Zimbabwe	10704340
1701	Zimbabwe	11404948
1702	Zimbabwe	11926563
1703	Zimbabwe	12311143

1704 rows × 2 columns

```
np.unique(df["country"], return_counts=True)
```

```
(array(['Afghanistan', 'Albania', 'Algeria', 'Angola', 'Argentina',
        'Australia', 'Austria', 'Bahrain', 'Bangladesh', 'Belgium',
        'Benin', 'Bolivia', 'Bosnia and Herzegovina', 'Botswana', 'Brazil',
        'Bulgaria', 'Burkina Faso', 'Burundi', 'Cambodia', 'Cameroon',
        'Canada', 'Central African Republic', 'Chad', 'Chile', 'China',
        'Colombia', 'Comoros', 'Congo, Dem. Rep.', 'Congo, Rep.',
        'Costa Rica', "Cote d'Ivoire", 'Croatia', 'Cuba', 'Czech Republic',
        'Denmark', 'Djibouti', 'Dominican Republic', 'Ecuador', 'Egypt',
        'El Salvador', 'Equatorial Guinea', 'Eritrea', 'Ethiopia',
        'Finland', 'France', 'Gabon', 'Gambia', 'Germany', 'Ghana',
        'Greece', 'Guatemala', 'Guinea', 'Guinea-Bissau', 'Haiti',
        'Honduras', 'Hong Kong, China', 'Hungary', 'Iceland', 'India',
        'Indonesia', 'Iran', 'Iraq', 'Ireland', 'Israel', 'Italy',
        'Jamaica', 'Japan', 'Jordan', 'Kenya', 'Korea, Dem. Rep.',
        'Korea, Rep.', 'Kuwait', 'Lebanon', 'Lesotho', 'Liberia', 'Libya',
        'Madagascar', 'Malawi', 'Malaysia', 'Mali', 'Mauritania',
        'Mauritius', 'Mexico', 'Mongolia', 'Montenegro', 'Morocco',
        'Mozambique', 'Myanmar', 'Namibia', 'Nepal', 'Netherlands',
```

```
           'New Zealand', 'Nicaragua', 'Niger', 'Nigeria', 'Norway', 'Oman',
           'Pakistan', 'Panama', 'Paraguay', 'Peru', 'Philippines', 'Poland',
           'Portugal', 'Puerto Rico', 'Reunion', 'Romania', 'Rwanda',
           'Sao Tome and Principe', 'Saudi Arabia', 'Senegal', 'Serbia',
           'Sierra Leone', 'Singapore', 'Slovak Republic', 'Slovenia',
           'Somalia', 'South Africa', 'Spain', 'Sri Lanka', 'Sudan',
           'Swaziland', 'Sweden', 'Switzerland', 'Syria', 'Taiwan',
           'Tanzania', 'Thailand', 'Togo', 'Trinidad and Tobago', 'Tunisia',
           'Turkey', 'Uganda', 'United Kingdom', 'United States', 'Uruguay',
           'Venezuela', 'Vietnam', 'West Bank and Gaza', 'Yemen, Rep.',
           'Zambia', 'Zimbabwe'], dtype=object),
    array([12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12,
           12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12,
           12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12,
           12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12,
           12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12,
           12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12,
           12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12,
           12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12,
           12, 12, 12, 12, 12, 12]))
```

```
df["country"].unique()
```

Saving...

```
array(['Afghanistan', 'Albania', 'Algeria', 'Angola', 'Argentina',
       'Australia', 'Austria', 'Bahrain', 'Bangladesh', 'Belgium',
       'Bosnia and Herzegovina', 'Botswana', 'Brazil',
       Faso', 'Burundi', 'Cambodia', 'Cameroon',
       rican Republic', 'Chad', 'Chile', 'China',
       'Colombia', 'Comoros', 'Congo, Dem. Rep.', 'Congo, Rep.',
       'Costa Rica', "Cote d'Ivoire", 'Croatia', 'Cuba', 'Czech Republic',
       'Denmark', 'Djibouti', 'Dominican Republic', 'Ecuador', 'Egypt',
       'El Salvador', 'Equatorial Guinea', 'Eritrea', 'Ethiopia',
       'Finland', 'France', 'Gabon', 'Gambia', 'Germany', 'Ghana',
       'Greece', 'Guatemala', 'Guinea', 'Guinea-Bissau', 'Haiti',
       'Honduras', 'Hong Kong, China', 'Hungary', 'Iceland', 'India',
       'Indonesia', 'Iran', 'Iraq', 'Ireland', 'Israel', 'Italy',
       'Jamaica', 'Japan', 'Jordan', 'Kenya', 'Korea, Dem. Rep.',
       'Korea, Rep.', 'Kuwait', 'Lebanon', 'Lesotho', 'Liberia', 'Libya',
       'Madagascar', 'Malawi', 'Malaysia', 'Mali', 'Mauritania',
       'Mauritius', 'Mexico', 'Mongolia', 'Montenegro', 'Morocco',
       'Mozambique', 'Myanmar', 'Namibia', 'Nepal', 'Netherlands',
       'New Zealand', 'Nicaragua', 'Niger', 'Nigeria', 'Norway', 'Oman',
       'Pakistan', 'Panama', 'Paraguay', 'Peru', 'Philippines', 'Poland',
       'Portugal', 'Puerto Rico', 'Reunion', 'Romania', 'Rwanda',
       'Sao Tome and Principe', 'Saudi Arabia', 'Senegal', 'Serbia',
       'Sierra Leone', 'Singapore', 'Slovak Republic', 'Slovenia',
       'Somalia', 'South Africa', 'Spain', 'Sri Lanka', 'Sudan',
       'Swaziland', 'Sweden', 'Switzerland', 'Syria', 'Taiwan',
       'Tanzania', 'Thailand', 'Togo', 'Trinidad and Tobago', 'Tunisia',
       'Turkey', 'Uganda', 'United Kingdom', 'United States', 'Uruguay',
       'Venezuela', 'Vietnam', 'West Bank and Gaza', 'Yemen, Rep.',
       'Zambia', 'Zimbabwe'], dtype=object)
```

```
df["country"].value_counts()
```

```
Afghanistan          12
Pakistan             12
New Zealand          12
Nicaragua            12
Niger                12
                     ..
Eritrea              12
Equatorial Guinea    12
El Salvador          12
Egypt                12
Zimbabwe             12
Name: country, Length: 142, dtype: int64
```

```
df.rename({"population": "Population", "country":"Country" }, axis = 1)
```

	Country	year	Population	continent	life_exp	gdp_cap
0	Afghanistan	1952	8425333	Asia	28.801	779.445314
1	Afghanistan	1957	9240934	Asia	30.332	820.853030
2	Afghanistan	1962	10267083	Asia	31.997	853.100710

```
df.rename(columns={"country":"Country"}) # wont suggest
```

	Country	year	population	continent	life_exp	gdp_cap
0	Afghanistan	1952	8425333	Asia	28.801	779.445314
1	Afghanistan	1957	9240934	Asia	30.332	820.853030
2	Afghanistan	1962	10267083	Asia	31.997	853.100710
3	Afghanistan	1967	11537966	Asia	34.020	836.197138
4	Afghanistan	1972	13079460	Asia	36.088	739.981106
...	...	...	...	...	...	...
1699	Zimbabwe	1987	9216418	Africa	62.351	706.157306
1700	Zimbabwe	1992	10704340	Africa	60.377	693.420786
			04948	Africa	46.809	792.449960
			26563	Africa	39.989	672.038623
1703	Zimbabwe	2007	12311143	Africa	43.487	469.709298

1704 rows × 6 columns


```
df.rename({"country": "Country"}, axis = 1, inplace = True)
df
```

	Country	year	population	continent	life_exp	gdp_cap
0	Afghanistan	1952	8425333	Asia	28.801	779.445314
1	Afghanistan	1957	9240934	Asia	30.332	820.853030
2	Afghanistan	1962	10267083	Asia	31.997	853.100710
3	Afghanistan	1967	11537966	Asia	34.020	836.197138
4	Afghanistan	1972	13079460	Asia	36.088	739.981106
...	...	...	...	...	...	...
1699	Zimbabwe	1987	9216418	Africa	62.351	706.157306
1700	Zimbabwe	1992	10704340	Africa	60.377	693.420786
1701	Zimbabwe	1997	11404948	Africa	46.809	792.449960
1702	Zimbabwe	2002	11926563	Africa	39.989	672.038623
1703	Zimbabwe	2007	12311143	Africa	43.487	469.709298

1704 rows × 6 columns

```
df
```

	Country	year	population	continent	life_exp	gdp_cap
0	Afghanistan	1952	8425333	Asia	28.801	779.445314
1	Afghanistan	1957	9240934	Asia	30.332	820.853030
2	Afghanistan	1962	10267083	Asia	31.997	853.100710
3	Afghanistan	1967	11537966	Asia	34.020	836.197138
4	Afghanistan	1972	13079460	Asia	36.088	739.981106
...	...	...	...	...	...	...
1699	Zimbabwe	1987	9216418	Africa	62.351	706.157306
1700	Zimbabwe	1992	10704340	Africa	60.377	693.420786
1701	Zimbabwe	1997	11404948	Africa	46.809	792.449960
1702	Zimbabwe	2002	11926563	Africa	39.989	672.038623
1703	Zimbabwe	2007	12311143	Africa	43.487	469.709298

1704 rows × 6 columns

```
df["Country"]
```

```
0       Afghanistan
1       Afghanistan
2       Afghanistan
3       Afghanistan
4       Afghanistan
           ...
1699       Zimbabwe
1700       Zimbabwe
1701       Zimbabwe
1702       Zimbabwe
1703       Zimbabwe
Name: Country, Length: 1704, dtype: object
```

```
df.Country # SERIOUSLY not recommended
# homework
# column name --> shape, shape is also an attribute to extract the shape
```

```
0       Afghanistan
1       Afghanistan
2       Afghanistan
3       Afghanistan
4       Afghanistan
```

Saving...

```
1701       Zimbabwe
1702       Zimbabwe
1703       Zimbabwe
Name: Country, Length: 1704, dtype: object
```

```
df.drop("continent", axis=1)
```

	Country	year	population	life_exp	gdp_cap
0	Afghanistan	1952	8425333	28.801	779.445314
1	Afghanistan	1957	9240934	30.332	820.853030
2	Afghanistan	1962	10267083	31.997	853.100710
3	Afghanistan	1967	11537966	34.020	836.197138
4	Afghanistan	1972	13079460	36.088	739.981106
...	...	...	...	...	...
1699	Zimbabwe	1987	9216418	62.351	706.157306
1700	Zimbabwe	1992	10704340	60.377	693.420786
1701	Zimbabwe	1997	11404948	46.809	792.449960
1702	Zimbabwe	2002	11926563	39.989	672.038623
1703	Zimbabwe	2007	12311143	43.487	469.709298

1704 rows × 5 columns

```
df["year+7"] = df["year"] + 7
```

```
df
```

```
df["gdp"] = df["gdp_cap"] * df["population"]
```

```
df
```

	Country	year	population	continent	life_exp	gdp_cap	year+7	gdp
0	Afghanistan	1952	8425333	Asia	28.801	779.445314	1959	6.567086e+09
1	Afghanistan	1957	9240934	Asia	30.332	820.853030	1964	7.585449e+09
2	Afghanistan	1962	10267083	Asia	31.997	853.100710	1969	8.758856e+09
3	Afghanistan	1967	11537966	Asia	34.020	836.197138	1974	9.648014e+09
4	Afghanistan	1972	13079460	Asia	36.088	739.981106	1979	9.678553e+09
...	...	...	...	...	...	...	...	...
1699	Zimbabwe	1987	9216418	Africa	62.351	706.157306	1994	6.508241e+09
1700	Zimbabwe	1992	10704340	Africa	60.377	693.420786	1999	7.422612e+09
1701	Zimbabwe	1997	11404948	Africa	46.809	792.449960	2004	9.037851e+09
1702	Zimbabwe	2002	11926563	Africa	39.989	672.038623	2009	8.015111e+09
				Africa	43.487	469.709298	2014	5.782658e+09

1704 rows × 8 columns


```
df["Own"] = [i for i in range(1704)]
df
```

	Country	year	population	continent	life_exp	gdp_cap	year+7	gdp	Own
0	Afghanistan	1952	8425333	Asia	28.801	779.445314	1959	6.567086e+09	0
1	Afghanistan	1957	9240934	Asia	30.332	820.853030	1964	7.585449e+09	1
2	Afghanistan	1962	10267083	Asia	31.997	853.100710	1969	8.758856e+09	2
3	Afghanistan	1967	11537966	Asia	34.020	836.197138	1974	9.648014e+09	3
4	Afghanistan	1972	13079460	Asia	36.088	739.981106	1979	9.678553e+09	4
...	...	...	...	...	...	...	...	...	...
1699	Zimbabwe	1987	9216418	Africa	62.351	706.157306	1994	6.508241e+09	1699
1700	Zimbabwe	1992	10704340	Africa	60.377	693.420786	1999	7.422612e+09	1700
1701	Zimbabwe	1997	11404948	Africa	46.809	792.449960	2004	9.037851e+09	1701
1702	Zimbabwe	2002	11926563	Africa	39.989	672.038623	2009	8.015111e+09	1702
1703	Zimbabwe	2007	12311143	Africa	43.487	469.709298	2014	5.782658e+09	1703

1704 rows × 9 columns

```
df.drop(["Own", "gdp", "year+7"], axis=1, inplace=True)
```

```
df
```

	Country	year	population	continent	life_exp	gdp_cap
1	Afghanistan	1957	9240934	Asia	30.332	820.853030
2	Afghanistan	1962	10267083	Asia	31.997	853.100710
3	Afghanistan	1967	11537966	Asia	34.020	836.197138
4	Afghanistan	1972	13079460	Asia	36.088	739.981106
...	...	...	...	...	...	...
1699	Zimbabwe	1987	9216418	Africa	62.351	706.157306
1700	Zimbabwe	1992	10704340	Africa	60.377	693.420786
1701	Zimbabwe	1997	11404948	Africa	46.809	792.449960
1702	Zimbabwe	2002	11926563	Africa	39.989	672.038623
1703	Zimbabwe	2007	12311143	Africa	43.487	469.709298

1704 rows × 6 columns

Saving...

0s completed at 23:01