Pandas

This section is all about the package, Pandas. Are you ready? Let's go!

You will find some small tasks in sections below. Most codes are hidden and you can only see the output.

> Try to figure out by yourself, or search for references. Being able to search and find information needed is an important skill that benefits you and your career for a long time.

Set up the environment

Task: import pandas, and name it as pd

```
import pandas as pd
```

Data Creation

Task

Find the data here for the GDP (nomial). For the top **10** countries, create a dataframe that has the names of country, the GDP of the country in 2021 (reported in 2022) by IMF in dollars , the population of the country (you can find the data here)

```
df = pd.DataFrame({'Country':['United States', 'China', 'Japan', 'Germany', 'India
                   'GDP': [25346805, 19911593, 4912147, 4256540, 3534743, 3376003,
                   'Population':[332943701, 1425881285, 125502000, 83695430, 14179
```

```
df
```

	Country	GDP	Population
0	United States	25346805	332943701
1	China	19911593	1425881285
2	Japan	4912147	125502000
3	Germany	4256540	83695430
4	India	3534743	1417945080
5	United Kingdom	3376003	67081234
6	France	2936702	67874000
7	Canada	2221218	38856839
8	Italy	2058330	58906742
9	Brazil	1833274	214956683

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
Data columns (total 3 columns):
 #   Column      Non-Null Count  Dtype 
---  ------      --------------  ----- 
 0   Country     10 non-null     object
 1   GDP         10 non-null     int64 
 2   Population  10 non-null     int64 
dtypes: int64(2), object(1)
memory usage: 368.0+ bytes
```

Data Accessing

Task: List the countries and their GDP

```
df[['Country', 'GDP']]
```

	Country	GDP
0	United States	25346805
1	China	19911593
2	Japan	4912147
3	Germany	4256540
4	India	3534743
5	United Kingdom	3376003
6	France	2936702
7	Canada	2221218
8	Italy	2058330
9	Brazil	1833274

Task: List the countries and their population

```
df[['Country', 'Population']]
```

	Country	Population
0	United States	332943701
1	China	1425881285
2	Japan	125502000
3	Germany	83695430
4	India	1417945080
5	United Kingdom	67081234
6	France	67874000
7	Canada	38856839
8	Italy	58906742
9	Brazil	214956683

Task: List the first country with its GDP and population

```
df.loc[0]
```

```
Country       United States
GDP                25346805
Population        332943701
Name: 0, dtype: object
```

Data Selection

Task: Show the countries whose GDP is more than 5 trillion dollars

> (Hint: the data source has US$ Million as a unit, hence you need to choose GDP > 5,000,000 unit)

```
df[df['GDP'] > 5000000]
```

	Country	GDP	Population
0	United States	25346805	332943701
1	China	19911593	1425881285

Task: Show the countries whose GDP is more than 3 trillion dollars

```
df[df['GDP'] > 3000000]
```

	Country	GDP	Population
0	United States	25346805	332943701
1	China	19911593	1425881285
2	Japan	4912147	125502000
3	Germany	4256540	83695430
4	India	3534743	1417945080
5	United Kingdom	3376003	67081234

Task: Show the countries whose pupulation is more than 1 billion

```
df[df['Population'] > 1000000000]
```

	Country	GDP	Population
1	China	19911593	1425881285
4	India	3534743	1417945080

Task: Show the countries whose pupulation is less than 300 million

```
df[df['Population'] < 300000000]
```

	Country	GDP	Population
2	Japan	4912147	125502000
3	Germany	4256540	83695430
5	United Kingdom	3376003	67081234
6	France	2936702	67874000
7	Canada	2221218	38856839
8	Italy	2058330	58906742
9	Brazil	1833274	214956683

Task: Show the countries whose pupulation is between 300 million and 500 million

```
df[(df['Population'] > 300000000) & (df['Population'] < 500000000)]
```

	Country	GDP	Population
0	United States	25346805	332943701

Data Manipulation

Task: Create a new column, called "GDP per capita", based on the column "GDP" and the column "Population"

```
df['GDP per capita'] = df['GDP'] / df['Population']
df
```

	Country	GDP	Population	GDP per capita
0	United States	25346805	332943701	0.076129
1	China	19911593	1425881285	0.013964
2	Japan	4912147	125502000	0.039140
3	Germany	4256540	83695430	0.050857
4	India	3534743	1417945080	0.002493
5	United Kingdom	3376003	67081234	0.050327
6	France	2936702	67874000	0.043267
7	Canada	2221218	38856839	0.057164
8	Italy	2058330	58906742	0.034942
9	Brazil	1833274	214956683	0.008529

Data Understanding

Task: With the three numeric columns, show the statsitics of each:

1. Count
2. Max
3. Min
4. Mean
5. Median
6. Quantiles
7. 25% Quantile
8. 50% Quantile
9. 75% QUantile
10. Variance
11. Std
12. Total

```
df.describe()
```

	GDP	Population	GDP per capita
count	1.000000e+01	1.000000e+01	10.000000
mean	7.038736e+06	3.833643e+08	0.037681
std	8.371837e+06	5.545693e+08	0.023292
min	1.833274e+06	3.885684e+07	0.002493
25%	2.400089e+06	6.727943e+07	0.019209
50%	3.455373e+06	1.045987e+08	0.041203
75%	4.748245e+06	3.034469e+08	0.050725
max	2.534680e+07	1.425881e+09	0.076129

```
df.median()
```

```
GDP               3.455373e+06
Population        1.045987e+08
GDP per capita    4.120348e-02
dtype: float64
```

```
df.var()
```

```
GDP               7.008765e+13
Population        3.075471e+17
GDP per capita    5.425327e-04
dtype: float64
```

```
df.sum()
```

```
Country           United StatesChinaJapanGermanyIndiaUnited King...
GDP                                                        70387355
Population                                               3833642994
GDP per capita                                             0.376813
dtype: object
```

Colab paid products - Cancel contracts here