```python
import pandas as pd
import os
import plotly.express as pe
import plotly.offline as pyo

df = pd.read_csv("gdp.csv")

df.head()

  Country Name Country Code  Year        Value
0   Arab World          ARB  1968  2.576068e+10
1   Arab World          ARB  1969  2.843420e+10
2   Arab World          ARB  1970  3.138550e+10
3   Arab World          ARB  1971  3.642691e+10
4   Arab World          ARB  1972  4.331606e+10

df.isnull().sum() # we had checked in this nothing is null

Country Name    0
Country Code    0
Year            0
Value           0
dtype: int64

df["Country Name"].duplicated().sum()

11251
```

Check description of each column

```python
df["Country Name"].describe()
# for hongkong sar , china we have more years of data that shows of 57
years
# similarly we can go for other categories

count                    11507
unique                     256
top        Hong Kong SAR, China
freq                        57
Name: Country Name, dtype: object

df["Year"].describe()
# shows we have minimum data of 1960 and maximum of 2016

count    11507.000000
mean      1991.265230
std         15.886648
min       1960.000000
25%       1978.000000
50%       1993.000000
```
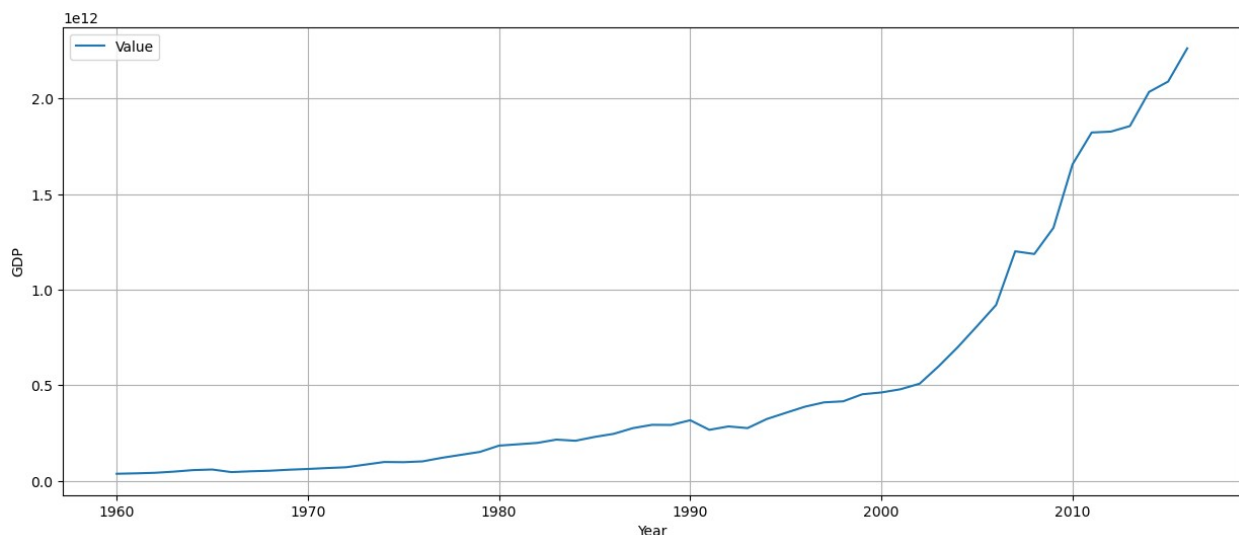
```
75%        2005.000000
max        2016.000000
Name: Year, dtype: float64
```

GDP Growth comparison between specific countries

```python
df_india = df[df["Country Name"]=="India"]
# here by changing the country name u can do analysis of that country

df_india.plot(kind="line"  , x = "Year" , y = "Value" , figsize =
(15,6) , grid =True , ylabel = "GDP")

<Axes: xlabel='Year', ylabel='GDP'>
```



```python
# to calculate gdp growth in 1969 we have to compare the data of 1969-
1968
round(((2843-2576)/2576)*100,2)
# gives that 10.36 % gdp increase

10.36

#On mass scale
# for that we have to convert it firstly into list

data = df_india.values
gdp_growth =[0]
for i in range(1,len(data)):

    prev = data[i-1][3]
    current =data[i][3]
    gdp_growth.append(round(((current-prev)/prev)*100,2))
    # now this where we get part to test gdp growth
    # this gives growth per year
```

```
    # so from here u can make great analysis on new column using
diffent graph
```

Finding GDP_GROWTH of a country

```
# this was the whole code for finding GDP_GROWTH of individual country
#df_india = df[df["Country Name"]=="India"]
# data = df_india.values
# gdp_growth =[0]
# for i in range(1,len(data)):

#      prev = data[i-1][3]
#      current =data[i][3]
#      gdp_growth.append(round(((current-prev)/prev)*100,2))

df_india.assign(GDP_GROWTH =gdp_growth)
```

	Country Name	Country Code	Year	Value	GDP_GROWTH
6074	India	IND	1960	3.653593e+10	0.00
6075	India	IND	1961	3.870910e+10	5.95
6076	India	IND	1962	4.159907e+10	7.47
6077	India	IND	1963	4.777600e+10	14.85
6078	India	IND	1964	5.572687e+10	16.64
6079	India	IND	1965	5.876042e+10	5.44
6080	India	IND	1966	4.525364e+10	-22.99
6081	India	IND	1967	4.946617e+10	9.31
6082	India	IND	1968	5.237732e+10	5.89
6083	India	IND	1969	5.766833e+10	10.10
6084	India	IND	1970	6.158980e+10	6.80
6085	India	IND	1971	6.645256e+10	7.90
6086	India	IND	1972	7.050991e+10	6.11
6087	India	IND	1973	8.437454e+10	19.66
6088	India	IND	1974	9.819828e+10	16.38
6089	India	IND	1975	9.715922e+10	-1.06
6090	India	IND	1976	1.013470e+11	4.31
6091	India	IND	1977	1.198667e+11	18.27
6092	India	IND	1978	1.354688e+11	13.02
6093	India	IND	1979	1.509508e+11	11.43
6094	India	IND	1980	1.838399e+11	21.79
6095	India	IND	1981	1.909095e+11	3.85
6096	India	IND	1982	1.980377e+11	3.73
6097	India	IND	1983	2.153508e+11	8.74
6098	India	IND	1984	2.093282e+11	-2.80
6099	India	IND	1985	2.294103e+11	9.59
6100	India	IND	1986	2.456647e+11	7.09
6101	India	IND	1987	2.753114e+11	12.07
6102	India	IND	1988	2.926327e+11	6.29
6103	India	IND	1989	2.920933e+11	-0.18

```
6104        India         IND   1990   3.166973e+11         8.42
6105        India         IND   1991   2.665023e+11       -15.85
6106        India         IND   1992   2.843639e+11         6.70
6107        India         IND   1993   2.755704e+11        -3.09
6108        India         IND   1994   3.229099e+11        17.18
6109        India         IND   1995   3.554760e+11        10.09
6110        India         IND   1996   3.876560e+11         9.05
6111        India         IND   1997   4.103203e+11         5.85
6112        India         IND   1998   4.157309e+11         1.32
6113        India         IND   1999   4.527000e+11         8.89
6114        India         IND   2000   4.621468e+11         2.09
6115        India         IND   2001   4.789655e+11         3.64
6116        India         IND   2002   5.080690e+11         6.08
6117        India         IND   2003   5.995929e+11        18.01
6118        India         IND   2004   6.996889e+11        16.69
6119        India         IND   2005   8.089011e+11        15.61
6120        India         IND   2006   9.203165e+11        13.77
6121        India         IND   2007   1.201112e+12        30.51
6122        India         IND   2008   1.186953e+12        -1.18
6123        India         IND   2009   1.323940e+12        11.54
6124        India         IND   2010   1.656617e+12        25.13
6125        India         IND   2011   1.823050e+12        10.05
6126        India         IND   2012   1.827638e+12         0.25
6127        India         IND   2013   1.856722e+12         1.59
6128        India         IND   2014   2.035393e+12         9.62
6129        India         IND   2015   2.089865e+12         2.68
6130        India         IND   2016   2.263792e+12         8.32
```

Finding GDP_GROWTH of every country making automation

```python
# to find for all countries we will automate in for loop
final_data=[]
for country_name in df["Country Name"].unique():
    df_all = df[df["Country Name"]==country_name]
    data = df_all.values
    gdp_growth_all =[0]
    for i in range(1,len(data)):
        prev = data[i-1][3]
        current =data[i][3]
        gdp_growth_all.append(round(((current-prev)/prev)*100,2))

    df_all = df_all.assign(GDP_GROWTH =gdp_growth_all)
    final_data.append(df_all)
```

```
final_data
#now we have to add it on different column in dataframe
df = pd.concat(final_data, axis=0)

df.head(59)
```

	Country Name	Country Code	Year	Value	GDP_GROWTH
0	Arab World	ARB	1968	2.576068e+10	0.00
1	Arab World	ARB	1969	2.843420e+10	10.38
2	Arab World	ARB	1970	3.138550e+10	10.38
3	Arab World	ARB	1971	3.642691e+10	16.06
4	Arab World	ARB	1972	4.331606e+10	18.91
5	Arab World	ARB	1973	5.501839e+10	27.02
6	Arab World	ARB	1974	1.051458e+11	91.11
7	Arab World	ARB	1975	1.163370e+11	10.64
8	Arab World	ARB	1976	1.448462e+11	24.51
9	Arab World	ARB	1977	1.673083e+11	15.51
10	Arab World	ARB	1978	1.835555e+11	9.71
11	Arab World	ARB	1979	2.486462e+11	35.46
12	Arab World	ARB	1980	3.381775e+11	36.01
13	Arab World	ARB	1981	3.485928e+11	3.08
14	Arab World	ARB	1982	3.243288e+11	-6.96
15	Arab World	ARB	1983	3.039625e+11	-6.28
16	Arab World	ARB	1984	3.079408e+11	1.31
17	Arab World	ARB	1985	3.038936e+11	-1.31
18	Arab World	ARB	1986	2.890292e+11	-4.89
19	Arab World	ARB	1987	3.126817e+11	8.18
20	Arab World	ARB	1988	3.075030e+11	-1.66

21	Arab World	ARB	1989	3.223251e+11	
4.82					
22	Arab World	ARB	1990	4.468772e+11	
38.64					
23	Arab World	ARB	1991	4.397792e+11	-
1.59					
24	Arab World	ARB	1992	4.711635e+11	
7.14					
25	Arab World	ARB	1993	4.765136e+11	
1.14					
26	Arab World	ARB	1994	4.875269e+11	
2.31					
27	Arab World	ARB	1995	5.237599e+11	
7.43					
28	Arab World	ARB	1996	5.782313e+11	
10.40					
29	Arab World	ARB	1997	6.132795e+11	
6.06					
30	Arab World	ARB	1998	5.915256e+11	-
3.55					
31	Arab World	ARB	1999	6.438897e+11	
8.85					
32	Arab World	ARB	2000	7.350251e+11	
14.15					
33	Arab World	ARB	2001	7.232828e+11	-
1.60					
34	Arab World	ARB	2002	7.290517e+11	
0.80					
35	Arab World	ARB	2003	8.231105e+11	
12.90					
36	Arab World	ARB	2004	9.638623e+11	
17.10					
37	Arab World	ARB	2005	1.184662e+12	
22.91					
38	Arab World	ARB	2006	1.404114e+12	
18.52					
39	Arab World	ARB	2007	1.637573e+12	
16.63					
40	Arab World	ARB	2008	2.078116e+12	
26.90					
41	Arab World	ARB	2009	1.795820e+12	-
13.58					
42	Arab World	ARB	2010	2.109646e+12	
17.48					
43	Arab World	ARB	2011	2.501554e+12	
18.58					
44	Arab World	ARB	2012	2.741239e+12	
9.58					
45	Arab World	ARB	2013	2.839627e+12	

```
3.59
46            Arab World          ARB  2014  2.906616e+12
2.36
47            Arab World          ARB  2015  2.563302e+12         -
11.81
48            Arab World          ARB  2016  2.504703e+12         -
2.29
49  Caribbean small states        CSS  1960  2.004785e+09
0.00
50  Caribbean small states        CSS  1961  2.169733e+09
8.23
51  Caribbean small states        CSS  1962  2.289495e+09
5.52
52  Caribbean small states        CSS  1963  2.431592e+09
6.21
53  Caribbean small states        CSS  1964  2.626896e+09
8.03
54  Caribbean small states        CSS  1965  2.828615e+09
7.68
55  Caribbean small states        CSS  1966  3.067844e+09
8.46
56  Caribbean small states        CSS  1967  3.293145e+09
7.34
57  Caribbean small states        CSS  1968  3.274646e+09         -
0.56
58  Caribbean small states        CSS  1969  3.563688e+09
8.83
```

```python
df.groupby["Country Name"].max()["GDP"]
# dont know why this is  not working
# this will give u average growth of country per year using
sort_values u can also sort the values
```

```
---------------------------------------------------------------------
-----
TypeError                                 Traceback (most recent call
last)
Cell In[22], line 1
----> 1 df.groupby["Country Name"].max()["GDP"]

TypeError: 'method' object is not subscriptable
```

```python
df_world = df[df["Country Name"]=="World"]
```

```python
df_world.head()
```

```
      Country Name Country Code  Year         Value  GDP_GROWTH
2249         World          WLD  1960  1.366678e+12        0.00
2250         World          WLD  1961  1.421788e+12        4.03
```

```
2251        World        WLD   1962   1.526955e+12          7.40
2252        World        WLD   1963   1.643752e+12          7.65
2253        World        WLD   1964   1.800796e+12          9.55
```

```python
fig =pe.line(df_world , x='Year' , y ='Value' , title ="world gdp
analysis")
# u can also check for india in same flow as well

#by adding range attribute into it would be easy to analyze or compare
two economies

# so now we will learn that how we can save graph offline
pyo.plot(fig , filename ="world gdp analysis.html")
# this will open new saved html page
```

```
'world gdp analysis.html'
```

GDO of each country

```python
os.mkdir("GDP All")
for country_name in df["Country Name"].unique():
    df_world = df[df["Country Name"]==country_name]
    fig =pe.line(df_world , x='Year' , y ='Value' , title
=country_name+"GDP analysis")
    pyo.plot(fig , filename ="GDP ALL/"+country_name+'.html' ,
auto_open =False)
```

GDP of each country with respect to world (80 trillion)

```python
os.mkdir("GDP wrt world")
for country_name in df["Country Name"].unique():
    df_world = df[df["Country Name"]==country_name]
    fig =pe.line(df_world , x='Year' , y ='Value' , title
=country_name+"GDP analysis" , range_y=(0,8000000000000))
    pyo.plot(fig , filename ="GDP wrt world/"+country_name+'.html' ,
auto_open =False)
```

Compare GDP across countries

```python
fig =pe.line(df , x='Year' , y ='Value' , title =country_name+"GDP
analysis of all in single frame " , color='Country Name')
pyo.plot(fig , filename ="GDP_in_single_frame.html" )

'GDP_in_single_frame.html'
```

To compare the GDP of two Countries

```python
c1 = df[df["Country Name"]=="India"]
c2 = df[df["Country Name"]=="China"]
df_IC = pd.concat([c1,c2] , axis=0)
df_IC.sample(10)
```

```
      Country Name Country Code   Year        Value  GDP_GROWTH
4066         China          CHN   1980  1.911492e+11        7.22
6087         India          IND   1973  8.437454e+10       19.66
6105         India          IND   1991  2.665023e+11      -15.85
6111         India          IND   1997  4.103203e+11        5.85
4071         China          CHN   1985  3.094880e+11       19.06
6076         India          IND   1962  4.159907e+10        7.47
6075         India          IND   1961  3.870910e+10        5.95
4062         China          CHN   1976  1.539405e+11       -5.81
4084         China          CHN   1998  1.029043e+12        7.01
6104         India          IND   1990  3.166973e+11        8.42
```

```python
fig =pe.line(df_IC , x='Year' , y ='Value' , title =country_name+"GDP
analysis of India and china" , color='Country Name')
pyo.plot(fig , filename ="GDP_analysis_of_india_china.html" )
# similarly we can comapre gdp another two countries
# try china and world and do analysis

'GDP_analysis_of_india_china.html'

# single cell code
# c1 = df[df["Country Name"]=="India"]
# c2 = df[df["Country Name"]=="China"]
# df_IC = pd.concat([c1,c2] , axis=0)
# fig =pe.line(df_IC , x='Year' , y ='Value' , title
=country_name+"GDP analysis of India and china" , color='Country
Name')
# pyo.plot(fig , filename ="GDP_analysis_of_india_china.html" )
```

Automating the comparison of more than one countries GDP's

```python
lst =["IND", "ITA", "USA" , "CHN"]
def compare_gdp(lst , is_open):

    dfs = []
    for i in lst:
        dfs.append(df[df['Country Code'] == i])

        df_pr = pd.concat(dfs, axis = 0)

    fig = pe.line(df_pr, x = 'Year', y = 'Value', title = 'GDP
Comparison - ' + '|'.join(lst),
                  color = 'Country Name')
    pyo.plot(fig, filename =   'two_countries.html' )
    # at place of values we can try gdp

compare_gdp(["IND","ITA"] , False)
# i dont know why it is not working but it is an correct way

#same operation as u have done with value we can perform with gdp
growth per year we have to just use GDP at value rather than value
```

GDPgrowth in some interval of years (1960 – 2016)

```python
# this removes outliers of missing year of
dfs=[]
for country_name in df["Country Name"].unique():
    df_pr= df[df["Country Name"]==country_name]
    if (len(df_pr)==57):# this removes outliers of missing year of
coutry here we are setting condition to make verification more clear
        dfs.append(df_pr)
df_pr = pd.concat(dfs, axis=0)
# now we have dataframe with that countries as df_pr

len(dfs) # now we can see that contries with less count are removed to
make data more precize

120

df_pr
# so now u can plot the graphs using plotly as we used previously
```

	Country Name	Country Code	Year	Value
GDP_GROWTH				

```
49       Caribbean small states    CSS  1960  2.004785e+09
0.00
50       Caribbean small states    CSS  1961  2.169733e+09
8.23
51       Caribbean small states    CSS  1962  2.289495e+09
5.52
52       Caribbean small states    CSS  1963  2.431592e+09
6.21
53       Caribbean small states    CSS  1964  2.626896e+09
8.03
...                          ...    ...  ...           ...
...
11502                   Zimbabwe    ZWE  2012  1.424249e+10
17.72
11503                   Zimbabwe    ZWE  2013  1.545177e+10
8.49
11504                   Zimbabwe    ZWE  2014  1.589105e+10
2.84
11505                   Zimbabwe    ZWE  2015  1.630467e+10
2.60
11506                   Zimbabwe    ZWE  2016  1.661996e+10
1.93

[6840 rows x 5 columns]
```