

```
In [3]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib import style
from matplotlib import rcParams
%matplotlib inline

import warnings
warnings.filterwarnings("ignore")

In [2]: data_frame = pd.read_csv("WHO-COVID-19-global-data.csv")#read csv into pandas
data_frame
```

Out[2]:

	Date_reported	Country_code	Country	WHO_region	New_cases	Cumulative_cases	New_deaths	Cumulative_deaths
0	2020-01-03	AF	Afghanistan	EMRO	0	0	0	0
1	2020-01-04	AF	Afghanistan	EMRO	0	0	0	0
2	2020-01-05	AF	Afghanistan	EMRO	0	0	0	0
3	2020-01-06	AF	Afghanistan	EMRO	0	0	0	0
4	2020-01-07	AF	Afghanistan	EMRO	0	0	0	0
...
271597	2023-02-17	ZW	Zimbabwe	AFRO	0	263642	0	5662
271598	2023-02-18	ZW	Zimbabwe	AFRO	0	263642	0	5662
271599	2023-02-19	ZW	Zimbabwe	AFRO	0	263642	0	5662
271600	2023-02-20	ZW	Zimbabwe	AFRO	0	263642	0	5662
271601	2023-02-21	ZW	Zimbabwe	AFRO	0	263642	0	5662

271602 rows × 8 columns

```
In [3]: df1 = data_frame.drop(["Country_code","New_cases","Cumulative_cases","New_deaths","Country","Date_reported"], axis = 1)
df1 #dropping Countrycode,Newcases,Cumulativedeaths,Newdeaths,Country,Datereported

Out[3]:
```

	WHO_region	Cumulative_deaths
0	EMRO	0
1	EMRO	0
2	EMRO	0
3	EMRO	0
4	EMRO	0
...
271597	AFRO	5662
271598	AFRO	5662
271599	AFRO	5662
271600	AFRO	5662
271601	AFRO	5662

271602 rows × 2 columns

```
In [4]: df1.describe()

Out[4]:
```

	Cumulative_deaths
count	2.716020e+05
mean	1.585335e+04
std	6.889940e+04
min	0.000000e+00
25%	1.100000e+01
50%	3.020000e+02
75%	4.631000e+03
max	1.106783e+06

```
In [5]: c = df1.groupby(["WHO_region"]).sum()# sum in WHOregion
c

Out[5]:
```

WHO_region	Cumulative_deaths
AFRO	116099288
AMRO	1954456944
EMRO	237565648
EURO	1345730489
Other	13986
SEARO	517237022
WPRO	134697634



```
In [7]: data_frame

Out[7]:
```

	Date_reported	Country_code	Country	WHO_region	New_cases	Cumulative_cases	New_deaths	Cumulative_deaths
0	2020-01-03	AF	Afghanistan	EMRO	0	0	0	0
1	2020-01-04	AF	Afghanistan	EMRO	0	0	0	0
2	2020-01-05	AF	Afghanistan	EMRO	0	0	0	0
3	2020-01-06	AF	Afghanistan	EMRO	0	0	0	0
4	2020-01-07	AF	Afghanistan	EMRO	0	0	0	0
...
271597	2023-02-17	ZW	Zimbabwe	AFRO	0	263642	0	5662
271598	2023-02-18	ZW	Zimbabwe	AFRO	0	263642	0	5662
271599	2023-02-19	ZW	Zimbabwe	AFRO	0	263642	0	5662
271600	2023-02-20	ZW	Zimbabwe	AFRO	0	263642	0	5662
271601	2023-02-21	ZW	Zimbabwe	AFRO	0	263642	0	5662

271602 rows × 8 columns

```
In [8]: r = data_frame.drop(["Country_code", "New_cases", "New_deaths", "Country", "Date_reported", "Cumulative_deaths"], axis = 1)
r #dropping Countrycode,Newcases,Cumulativedeaths,Newdeaths,Country,Datereported

Out[8]:
```

	WHO_region	Cumulative_cases
0	EMRO	0
1	EMRO	0
2	EMRO	0
3	EMRO	0
4	EMRO	0
...
271597	AFRO	263642
271598	AFRO	263642
271599	AFRO	263642
271600	AFRO	263642
271601	AFRO	263642

271602 rows × 2 columns

```
In [9]: s = r.groupby(["WHO_region"]).sum()# sum in WHOregion
s

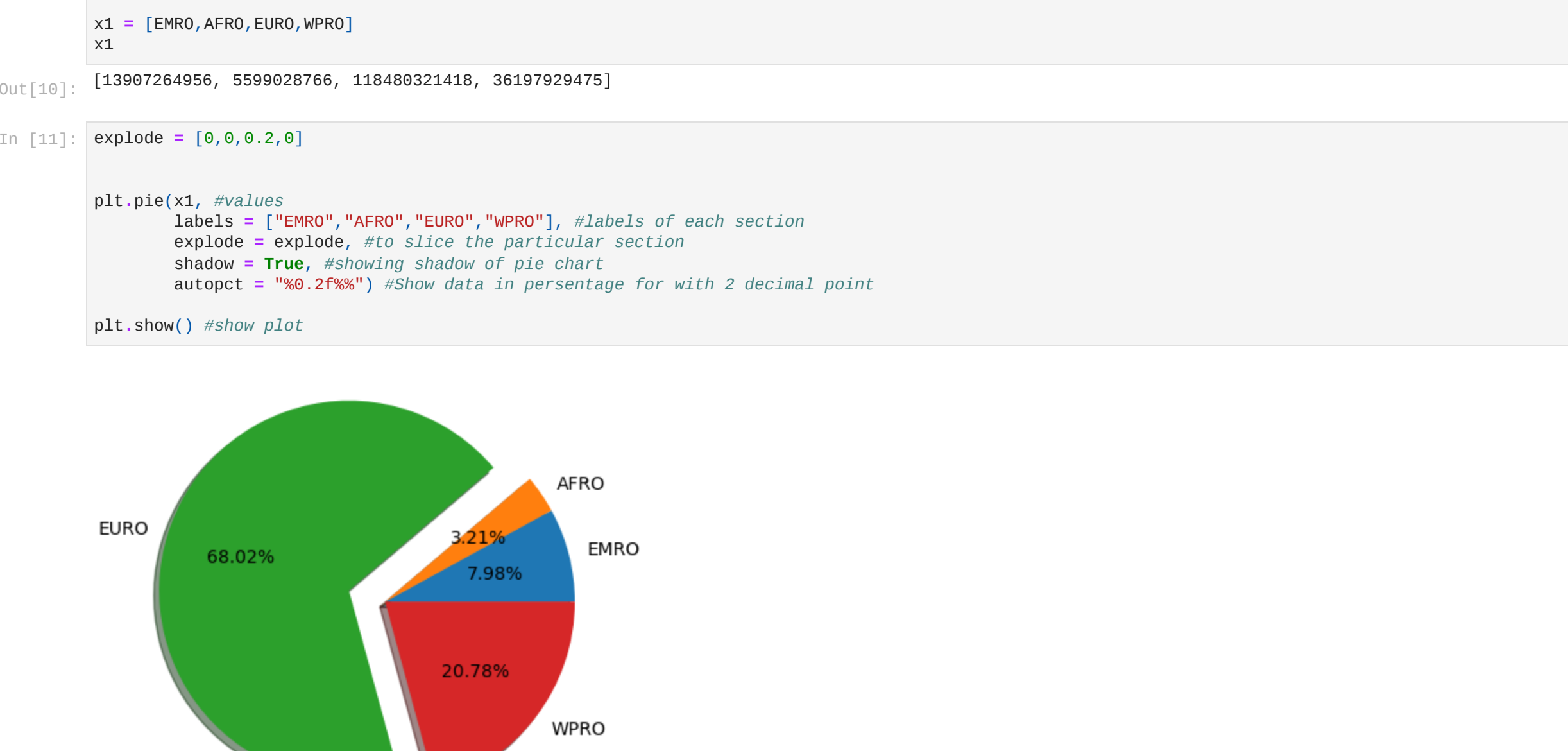
Out[9]:
```

WHO_region	Cumulative_cases
AFRO	5599028766
AMRO	98814961876
EMRO	13907264956
EURO	118480321418
Other	829560
SEARO	36769327691
WPRO	36197929475

```
In [10]: EMRO = r.loc[ r["WHO_region"] == "EMRO"]["Cumulative_cases"].sum()
AFRO = r.loc[ r["WHO_region"] == "AFRO"]["Cumulative_cases"].sum()
EURO = r.loc[ r["WHO_region"] == "EURO"]["Cumulative_cases"].sum()
WPRO = r.loc[ r["WHO_region"] == "WPRO"]["Cumulative_cases"].sum()

x1 = [EMRO, AFRO, EURO, WPRO]
x1

Out[10]: [13907264956, 5599028766, 118480321418, 36197929475]
```



```
In [12]: data_frame

Out[12]:
```

	Date_reported	Country_code	Country	WHO_region	New_cases	Cumulative_cases	New_deaths	Cumulative_deaths
0	2020-01-03	AF	Afghanistan	EMRO	0	0	0	0
1	2020-01-04	AF	Afghanistan	EMRO	0	0	0	0
2	2020-01-05	AF	Afghanistan	EMRO	0	0	0	0
3	2020-01-06	AF	Afghanistan	EMRO	0	0	0	0
4	2020-01-07	AF	Afghanistan	EMRO	0	0	0	0
...
271597	2023-02-17	ZW	Zimbabwe	AFRO	0	263642	0	5662
271598	2023-02-18	ZW	Zimbabwe	AFRO	0	263642	0	5662
271599	2023-02-19	ZW	Zimbabwe	AFRO	0	263642	0	5662
271600	2023-02-20	ZW	Zimbabwe	AFRO	0	263642	0	5662
271601	2023-02-21	ZW	Zimbabwe	AFRO	0	263642	0	5662

271602 rows × 8 columns

```
In [13]: t = data_frame.drop(["Country_code", "Country", "New_deaths", "Country", "Cumulative_cases", "Cumulative_deaths", "WHO_region"], axis = 1)
t #dropping Countrycode,Newdeaths,Cumulativedeaths,Newdeaths,Country,Datereported,Cumulative_cases

Out[13]:
```

	Date_reported	New_cases
0	2020-01-03	0
1	2020-01-04	0
2	2020-01-05	0
3	2020-01-06	0
4	2020-01-07	0
...
271597	2023-02-17	0
271598	2023-02-18	0
271599	2023-02-19	0
271600	2023-02-20	0
271601	2023-02-21	0

271602 rows × 2 columns

```
In [14]: datecol = pd.DatetimeIndex(t["Date_reported"]) #reading datetimeindex
t["year"] = datecol.year # year
t["month"] = datecol.month # month
t["day"] = datecol.day # day

In [15]: t

Out[15]:
```

	Date_reported	New_cases	year	month	day
0	2020-01-03	0	2020	1	3
1	2020-01-04	0	2020	1	4
2	2020-01-05	0	2020	1	5
3	2020-01-06	0	2020	1	6
4	2020-01-07	0	2020	1	7
...
271597	2023-02-17	0	2023	2	17
271598	2023-02-18	0	2023	2	18
271599	2023-02-19	0	2023	2	19
271600	2023-02-20	0	2023	2	20
271601	2023-02-21	0	2023	2	21

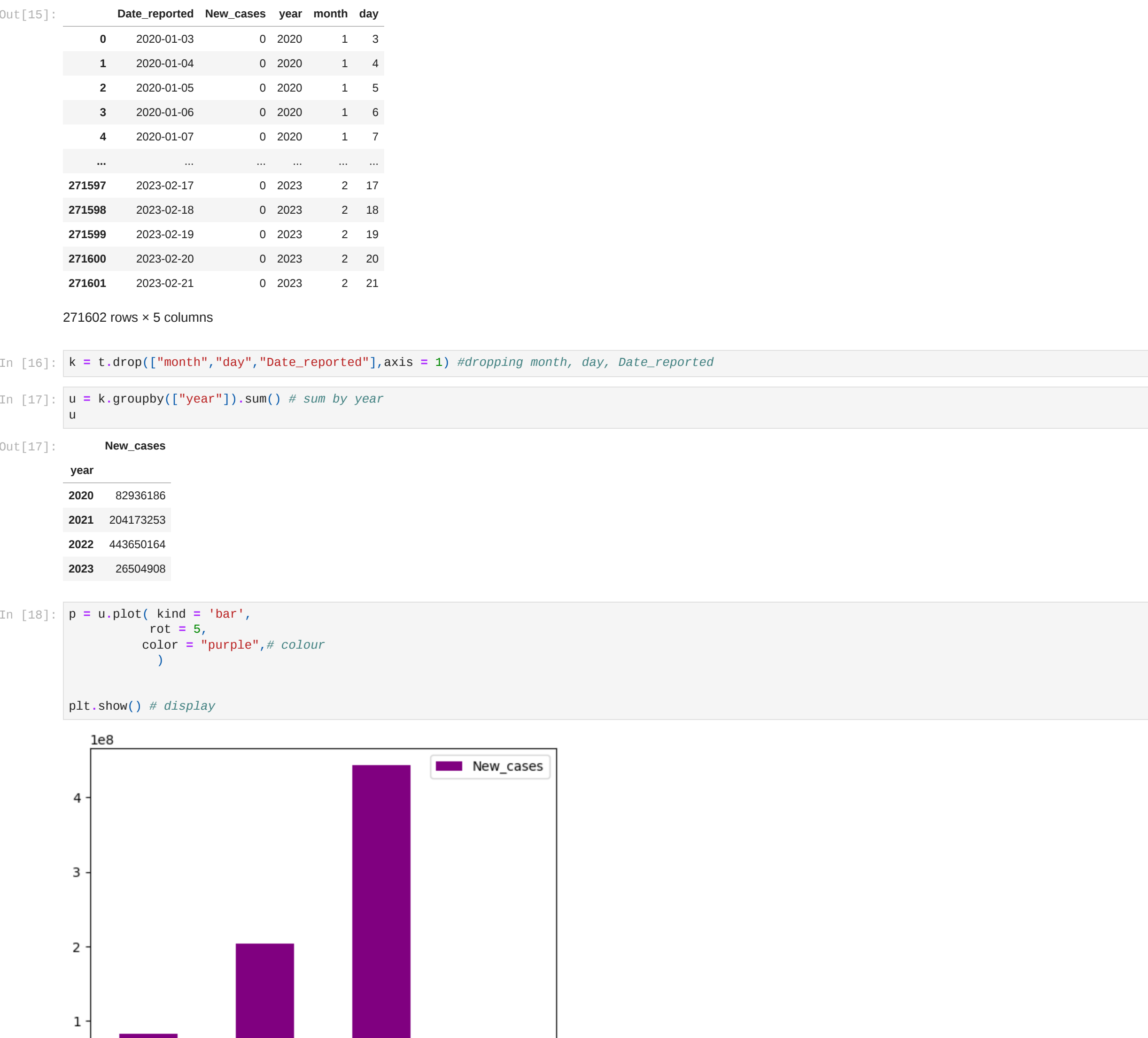
271602 rows × 5 columns

```
In [16]: k = t.drop(["month","day","Date_reported"],axis = 1) #dropping month, day, Date_reported

In [17]: u = k.groupby(["year"]).sum() # sum by year
u

Out[17]:
```

	New_cases
year	
2020	82936186
2021	204173253
2022	443650164
2023	26504908



```
In [ ]:
```