United States - Crime Rates - 1960 - 2014

Introduction:

This time you will create a data

Special thanks to: https://github.com/justmarkham (https://github.com/justmarkham) for sharing the dataset and materials.

Step 1. Import the necessary libraries

In [ ]:
```
import pandas as pd
import numpy as np

import datetime
```

Step 2. Import the dataset from this address (https://raw.githubusercontent.com/guipsamora/pandas_exercises/mast

Step 3. Assign it to a variable called crime.

In [ ]:
```
data_url = "https://raw.githubusercontent.com/a-forty-two/COG_GN22CDBDS002_11_1_
crime = pd.read_csv(data_url)
```

In [ ]:
```
crime.tail()
```

Out[134]:

	Year	Population	Total	Violent	Property	Murder	Forcible_Rape	Robbery	Aggravated_
50	2010	309330219	10363873	1251248	9112625	14772	85593	369089	
51	2011	311587816	10258774	1206031	9052743	14661	84175	354772	
52	2012	313873685	10219059	1217067	9001992	14866	85141	355051	
53	2013	316497531	9850445	1199684	8650761	14319	82109	345095	
54	2014	318857056	9475816	1197987	8277829	14249	84041	325802	

Step 4. What is the type of the columns?

In [ ]:
```
crime.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 55 entries, 0 to 54
Data columns (total 12 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   Year                55 non-null     int64
 1   Population          55 non-null     int64
 2   Total               55 non-null     int64
 3   Violent             55 non-null     int64
 4   Property            55 non-null     int64
 5   Murder              55 non-null     int64
 6   Forcible_Rape       55 non-null     int64
 7   Robbery             55 non-null     int64
 8   Aggravated_assault  55 non-null     int64
 9   Burglary            55 non-null     int64
 10  Larceny_Theft       55 non-null     int64
 11  Vehicle_Theft       55 non-null     int64
dtypes: int64(12)
memory usage: 5.3 KB
```

Have you noticed that the type of Year is int64. But pandas has a different type to work with Time Series. Let's see it now.

Step 5. Convert the type of the column Year to datetime64

In [ ]:
```
crime['Year'] = pd.to_datetime(crime['Year'], format='%Y')
```

In [ ]:
```
crime.tail()
```

Out[137]:

	Year	Population	Total	Violent	Property	Murder	Forcible_Rape	Robbery	Aggravated
50	2010-01-01	309330219	10363873	1251248	9112625	14772	85593	369089	
51	2011-01-01	311587816	10258774	1206031	9052743	14661	84175	354772	
52	2012-01-01	313873685	10219059	1217067	9001992	14866	85141	355051	
53	2013-01-01	316497531	9850445	1199684	8650761	14319	82109	345095	
54	2014-01-01	318857056	9475816	1197987	8277829	14249	84041	325802	

Step 6. Set the Year column as the index of the dataframe

In [ ]:
```
crime = crime.set_index('Year')
crime.head()
```

Out[138]:

Year	Population	Total	Violent	Property	Murder	Forcible_Rape	Robbery	Aggravated_assau
1960-01-01	179323175	3384200	288460	3095700	9110	17190	107840	15432
1961-01-01	182992000	3488000	289390	3198600	8740	17220	106670	15676
1962-01-01	185771000	3752200	301510	3450700	8530	17550	110860	16457
1963-01-01	188483000	4109500	316970	3792500	8640	17650	116470	17421
1964-01-01	191141000	4564600	364220	4200400	9360	21420	130390	20305

Step 7. Delete the Total column

In [ ]:
```
crime.drop(['Total'],axis =1,inplace =True);
```

Step 8. Group the year by decades and sum the values

Pay attention to the Population column number, summing this column is a mistake

In [ ]:
```
crime.groupby((crime.index.year//10)*10).sum()
```

Out[164]:

Year	Population	Violent	Property	Murder	Forcible_Rape	Robbery	Aggravated_assault	B
1960	1915053175	4134930	45160900	106180	236720	1633510	2158520	13
1970	2121193298	9607930	91383800	192230	554570	4159020	4702120	28
1980	2371370069	14074328	117048900	206439	865639	5383109	7619130	33
1990	2612825258	17527048	119053499	211664	998827	5748930	10568963	26
2000	2947969117	13968056	100944369	163068	922499	4230366	8652124	21
2010	1570146307	6072017	44095950	72867	421059	1749809	3764142	10

Step 9. What is the most dangerous decade to live in the US?

In [ ]:
```
crime.idxmax(0)
```

Out[165]:
```
Population           2014-01-01
Violent              1992-01-01
Property             1991-01-01
Murder               1991-01-01
Forcible_Rape        1992-01-01
Robbery              1991-01-01
Aggravated_assault   1993-01-01
Burglary             1980-01-01
Larceny_Theft        1991-01-01
Vehicle_Theft        1991-01-01
dtype: datetime64[ns]
```

In [ ]:
```
df = crime.groupby((crime.index.year//10)*10).sum()
df.drop(['Population','Property'],axis =1,inplace =True);
```

In [ ]:
```
df.sum(axis=1)
```

Out[273]:
```
Year
1960     53430760
1970    110599570
1980    145197803
1990    154109231
2000    128884414
2010     56175842
dtype: int64
```

In [ ]:
```
df.sum(axis=1).max()
```

Out[274]:
```
154109231
```

1990 is the most dangerous year to live in.