United States - Crime Rates - 1960 - 2014

Introduction:

This time you will create a data

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

Step 1. Import the necessary libraries

```
In [1]:    import numpy as np
           import pandas as pd
```

Step 2. Import the dataset from this address.

Step 3. Assign it to a variable called crime.

```
In [2]:    url = "https://raw.githubusercontent.com/guipsamora/pandas_exercises/master/04_Apply/US_Crime_Rates/US_Crime_Rates_1960_2014.csv"
           crime = pd.read_csv(url)
           crime.head()
```

Out[2]:

	Year	Population	Total	Violent	Property	Murder	Forcible_Rape	Robbery	Aggravated_assault	Burglary	Larceny_Theft	Vehicle_Theft
0	1960	179323175	3384200	288460	3095700	9110	17190	107840	154320	912100	1855400	328200
1	1961	182992000	3488000	289390	3198600	8740	17220	106670	156760	949600	1913000	336000
2	1962	185771000	3752200	301510	3450700	8530	17550	110860	164570	994300	2089600	366800
3	1963	188483000	4109500	316970	3792500	8640	17650	116470	174210	1086400	2297800	408300
4	1964	191141000	4564600	364220	4200400	9360	21420	130390	203050	1213200	2514400	472800

Step 4. What is the type of the columns?

```
In [3]:    crime.dtypes
```

```
Out[3]:    Year                  int64
           Population            int64
           Total                 int64
           Violent               int64
           Property              int64
           Murder                int64
           Forcible_Rape         int64
           Robbery               int64
           Aggravated_assault    int64
           Burglary              int64
           Larceny_Theft         int64
           Vehicle_Theft         int64
           dtype: object
```

```
In [4]:    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 [5]:    crime["Year"] = pd.to_datetime(crime["Year"], format="%Y")
           crime.head(1)
```

Out[5]:

	Year	Population	Total	Violent	Property	Murder	Forcible_Rape	Robbery	Aggravated_assault	Burglary	Larceny_Theft	Vehicle_Theft
0	1960-01-01	179323175	3384200	288460	3095700	9110	17190	107840	154320	912100	1855400	328200

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

```
In [6]:    crime.set_index("Year",inplace=True)
```

```
In [7]:    crime.head(1)
```

Out[7]:

Year	Population	Total	Violent	Property	Murder	Forcible_Rape	Robbery	Aggravated_assault	Burglary	Larceny_Theft	Vehicle_Theft
1960-01-01	179323175	3384200	288460	3095700	9110	17190	107840	154320	912100	1855400	328200

Step 7. Delete the Total column

```
In [8]:    del crime['Total']
           crime.head()
```

Out[8]:

Year	Population	Violent	Property	Murder	Forcible_Rape	Robbery	Aggravated_assault	Burglary	Larceny_Theft	Vehicle_Theft
1960-01-01	179323175	288460	3095700	9110	17190	107840	154320	912100	1855400	328200
1961-01-01	182992000	289390	3198600	8740	17220	106670	156760	949600	1913000	336000
1962-01-01	185771000	301510	3450700	8530	17550	110860	164570	994300	2089600	366800
1963-01-01	188483000	316970	3792500	8640	17650	116470	174210	1086400	2297800	408300
1964-01-01	191141000	364220	4200400	9360	21420	130390	203050	1213200	2514400	472800

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 [9]:    """
           Resampling is necessary when you're given a data set recorded in some time interval and you want to change the time interval
           to something else.
           For example, you could aggregate monthly data into yearly data, or you could upsample hourly data into minute-by-minute data.
           """

           # Uses resample to sum each decade
           crimes = crime.resample('10AS').sum()

           # Uses resample to get the max value only for the "Population" column
           population = crime['Population'].resample('10AS').max()

           # Updating the "Population" column
           crimes['Population'] = population

           crimes
```

Out[9]:

Year	Population	Violent	Property	Murder	Forcible_Rape	Robbery	Aggravated_assault	Burglary	Larceny_Theft	Vehicle_Theft
1960-01-01	201385000	4134930	45160900	106180	236720	1633510	2158520	13321100	26547700	5292100
1970-01-01	220099000	9607930	91383800	192230	554550	4159020	4702120	28486000	53157800	9739900
1980-01-01	248239000	14074328	117048900	206439	865639	5383109	7619130	33073494	72040253	11935411
1990-01-01	272690813	17527048	119053499	211664	998827	5748930	10568963	26750015	77679366	14624418
2000-01-01	307006550	13968056	100944369	163068	922499	4230366	8652124	21565176	67970291	11412834
2010-01-01	318857056	6072017	44095950	72867	421059	1749809	3764142	10125170	30401698	3569080

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

```
In [10]:   # apparently the 90s was a pretty dangerous time in the US
           crime.idxmax(0)
```

```
Out[10]:   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 [ ]:
```