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

```
import pandas as pd
```

Step 2. Import the dataset from this address.

Step 3. Assign it to a variable called crime.

```
crime = pd.read_csv('https://raw.githubusercontent.com/thieu1995/csv-files/main/data/pandas/US_Crime_Rates_1960_2014.csv')
crime.head()
```

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

Next steps: Generate code with crime | View recommended plots | New interactive sheet

Step 4. What is the type of the columns?

```
crime.dtypes
```

	0
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

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

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

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

```
crime.set_index('Year', inplace=True)
```

Step 7. Delete the Total column

```
crime_c = crime.copy()
del crime['Total']
crime_c.head()
```

Year	Population	Total	Violent	Property	Murder	Forcible_Rape	Robbery	Aggravated_assault	Burglary	Larceny_Theft	Vehicle_T
1960-01-01	179323175	3384200	288460	3095700	9110	17190	107840	154320	912100	1855400	32
1961-01-01	182992000	3488000	289390	3198600	8740	17220	106670	156760	949600	1913000	33
1962-01-01	185771000	3752200	301510	3450700	8530	17550	110860	164570	994300	2089600	36

Next steps: Generate code with crime_c | View recommended plots | New interactive sheet

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

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

```
crime['Decade'] = (crime.index.year // 10) * 10
crime_by_decade = crime.groupby('Decade').sum()
```

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

```
crime_by_decade['Total_Crime'] = crime_by_decade.sum(axis=1)
dangerous = crime_by_decade['Total_Crime'].idxmax()

print(f"the most dangerous decade: {dangerous}s.")
```

```
the most dangerous decade: 2000s.
```