

BigDataProject

December 11, 2021

Analysis of Police Officers Deaths in the United States

CSCI 5446

Team:

Pavan Kumar Kadambala

Syam Sai Krishna Kaduthuri

Sandeep Chowdary Kotapati

Before you begin completing the assignment, execute the following cell to load the provided tests.

1. Obtaining the data

```
[9]: import pandas as pd
      from datetime import datetime
      import numpy as np
```

```
[10]: !wget https://github.com/fivethirtyeight/data/raw/master/police-deaths/all_data.
      ↪ csv
```

```
--2021-12-07 22:23:14--
https://github.com/fivethirtyeight/data/raw/master/police-deaths/all_data.csv
Resolving github.com (github.com)... 140.82.113.3
Connecting to github.com (github.com)|140.82.113.3|:443... connected.
HTTP request sent, awaiting response... 302 Found
Location: https://raw.githubusercontent.com/fivethirtyeight/data/master/police-
deaths/all_data.csv [following]
--2021-12-07 22:23:14--
https://raw.githubusercontent.com/fivethirtyeight/data/master/police-
deaths/all_data.csv
Resolving raw.githubusercontent.com (raw.githubusercontent.com)...
185.199.110.133, 185.199.109.133, 185.199.108.133, ...
Connecting to raw.githubusercontent.com
(raw.githubusercontent.com)|185.199.110.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
```

Length: 2966647 (2.8M) [text/plain]

Saving to: 'all_data.csv.1'

all_data.csv.1 100%[=====>] 2.83M --.-KB/s in 0.03s

2021-12-07 22:23:17 (99.2 MB/s) - 'all_data.csv.1' saved [2966647/2966647]

2. Data Cleaning

```
[763]: import numpy as np
import pandas as pd
import os
from datetime import datetime
```

```
[858]: # ODMF : Officer Down Memorial Page
odmp_data = pd.read_csv('all_data.csv')
```

```
[859]: odmp_data
```

```
[859]:
```

	person \
0	Constable Darius Quimby
1	Sheriff Cornelius Hogeboom
2	Deputy Sheriff Isaac Smith
3	Marshal Robert Forsyth
4	Sheriff Robert Maxwell
...	...
22795	K9 Bruno
22796	K9 Lazer
22797	K9 Tyson
22798	K9 Credo
22799	K9 Roscoe

	dept \
0	Albany County Constable's Office, NY
1	Columbia County Sheriff's Office, NY
2	Westchester County Sheriff's Department, NY
3	United States Department of Justice - United S...
4	Greenville County Sheriff's Office, SC
...	...
22795	Amarillo Police Department, TX
22796	United States Department of Homeland Security ...
22797	Fountain County Sheriff's Office, IN
22798	Long Beach Police Department, CA
22799	Emmett Police Department, ID

	eow	cause
0	EOW: Monday, January 3, 1791	Cause of Death: Gunfire

1	EOW: Saturday, October 22, 1791	Cause of Death: Gunfire
2	EOW: Thursday, May 17, 1792	Cause of Death: Gunfire
3	EOW: Saturday, January 11, 1794	Cause of Death: Gunfire
4	EOW: Sunday, November 12, 1797	Cause of Death: Gunfire
...
22795	EOW: Sunday, June 12, 2016	Cause of Death: Accidental
22796	EOW: Monday, June 20, 2016	Cause of Death: Heat exhaustion
22797	EOW: Monday, June 27, 2016	Cause of Death: Heat exhaustion
22798	EOW: Tuesday, June 28, 2016	Cause of Death: Gunfire (Accidental)
22799	EOW: Friday, July 1, 2016	Cause of Death: Automobile accident

[22800 rows x 4 columns]

2.1 Removing Null and invalid fields

```
[860]: odmp_data.isnull()
```

```
[860]:
```

	person	dept	eow	cause
0	False	False	False	False
1	False	False	False	False
2	False	False	False	False
3	False	False	False	False
4	False	False	False	False
...
22795	False	False	False	False
22796	False	False	False	False
22797	False	False	False	False
22798	False	False	False	False
22799	False	False	False	False

[22800 rows x 4 columns]

```
[861]: missing_value= ["N/a", "na", np.nan]
df_remv = pd.read_csv("all_data.csv", na_values = missing_value)
```

```
[862]: df_remv
```

```
[862]:
```

	person \
0	Constable Darius Quimby
1	Sheriff Cornelius Hogeboom
2	Deputy Sheriff Isaac Smith
3	Marshal Robert Forsyth
4	Sheriff Robert Maxwell
...	...
22795	K9 Bruno
22796	K9 Lazer
22797	K9 Tyson

```

22798          K9 Credo
22799          K9 Roscoe

                                dept \
0          Albany County Constable's Office, NY
1          Columbia County Sheriff's Office, NY
2          Westchester County Sheriff's Department, NY
3          United States Department of Justice - United S...
4          Greenville County Sheriff's Office, SC
...
22795          Amarillo Police Department, TX
22796 United States Department of Homeland Security ...
22797          Fountain County Sheriff's Office, IN
22798          Long Beach Police Department, CA
22799          Emmett Police Department, ID

```

```

                                eow                                cause
0          EOW: Monday, January 3, 1791          Cause of Death: Gunfire
1          EOW: Saturday, October 22, 1791          Cause of Death: Gunfire
2          EOW: Thursday, May 17, 1792          Cause of Death: Gunfire
3          EOW: Saturday, January 11, 1794          Cause of Death: Gunfire
4          EOW: Sunday, November 12, 1797          Cause of Death: Gunfire
...
22795          EOW: Sunday, June 12, 2016          Cause of Death: Accidental
22796          EOW: Monday, June 20, 2016          Cause of Death: Heat exhaustion
22797          EOW: Monday, June 27, 2016          Cause of Death: Heat exhaustion
22798          EOW: Tuesday, June 28, 2016          Cause of Death: Gunfire (Accidental)
22799          EOW: Friday, July 1, 2016          Cause of Death: Automobile accident

```

[22800 rows x 4 columns]

2.2 Adding new fields to the data

Adding 'year' field

```
[816]: df_remv['year'] = df_remv['eow'].str.split(' ').str[4]
```

```
[817]: df_remv['year']
```

```

[817]: 0          1791
1          1791
2          1792
3          1794
4          1797
...
22795      2016
22796      2016
22797      2016

```

```

22798    2016
22799    2016
Name: year, Length: 22800, dtype: object

```

```
[818]: df_remv
```

```

[818]:
           person \
0      Constable Darius Quimby
1      Sheriff Cornelius Hogeboom
2      Deputy Sheriff Isaac Smith
3      Marshal Robert Forsyth
4      Sheriff Robert Maxwell
...
22795      K9 Bruno
22796      K9 Lazer
22797      K9 Tyson
22798      K9 Credo
22799      K9 Roscoe

           dept \
0      Albany County Constable's Office, NY
1      Columbia County Sheriff's Office, NY
2      Westchester County Sheriff's Department, NY
3      United States Department of Justice - United S...
4      Greenville County Sheriff's Office, SC
...
22795      Amarillo Police Department, TX
22796      United States Department of Homeland Security ...
22797      Fountain County Sheriff's Office, IN
22798      Long Beach Police Department, CA
22799      Emmett Police Department, ID

           eow           cause \
0      EOW: Monday, January 3, 1791      Cause of Death: Gunfire
1      EOW: Saturday, October 22, 1791      Cause of Death: Gunfire
2      EOW: Thursday, May 17, 1792      Cause of Death: Gunfire
3      EOW: Saturday, January 11, 1794      Cause of Death: Gunfire
4      EOW: Sunday, November 12, 1797      Cause of Death: Gunfire
...
22795      EOW: Sunday, June 12, 2016      Cause of Death: Accidental
22796      EOW: Monday, June 20, 2016      Cause of Death: Heat exhaustion
22797      EOW: Monday, June 27, 2016      Cause of Death: Heat exhaustion
22798      EOW: Tuesday, June 28, 2016      Cause of Death: Gunfire (Accidental)
22799      EOW: Friday, July 1, 2016      Cause of Death: Automobile accident

           year
0      1791

```

```

1      1791
2      1792
3      1794
4      1797
...
22795  2016
22796  2016
22797  2016
22798  2016
22799  2016

```

[22800 rows x 5 columns]

Adding 'date' & 'month' field

```

[819]: df_dtext= df_remv['eow'].str.split(',').str[1]
df_dtext = df_dtext.str.strip()
df_mon= df_dtext.str.split(' ').str[0]
df_day= df_dtext.str.split(' ').str[1]

```

```

[820]: df_month = df_mon.map({'January':'01', 'February' : '02' , 'March':'03', 'April':
    ↳ '04', 'May':'05', 'June':'06', 'July':'07' , 'August':'08',
    ↳ 'September':'09', 'October':'10', 'November':'11',
    ↳ 'December':'12'})
df_day = df_day.replace({'1':'01', '2' : '02' , '3':'03', '4':'04', '5':'05',
    ↳ '6':'06', '7':'07' , '8':'08', '9':'09'})

```

```

[821]: df_day

```

```

[821]: 0      03
1      22
2      17
3      11
4      12
...
22795  12
22796  20
22797  27
22798  28
22799  01
Name: eow, Length: 22800, dtype: object

```

```

[822]: #df_day=df_day.astype(int)
df_remv['year']=df_remv['year'].astype(int)

```

```

[823]: df_remv['date']=df_month.astype(str)+'-'+df_day.astype(str)+'-'+df_remv['year'] .
    ↳ astype(str)

```

```
df_remv['month']=df_mon
```

```
[824]: df_remv[['month']]
```

```
[824]:
```

	month
0	January
1	October
2	May
3	January
4	November
...	...
22795	June
22796	June
22797	June
22798	June
22799	July

[22800 rows x 1 columns]

```
[825]: df_remv[['date']]
```

```
[825]:
```

	date
0	01-03-1791
1	10-22-1791
2	05-17-1792
3	01-11-1794
4	11-12-1797
...	...
22795	06-12-2016
22796	06-20-2016
22797	06-27-2016
22798	06-28-2016
22799	07-01-2016

[22800 rows x 1 columns]

```
[826]: df_remv[['year']]
```

```
[826]:
```

	year
0	1791
1	1791
2	1792
3	1794
4	1797
...	...
22795	2016
22796	2016

```
22797 2016
22798 2016
22799 2016
```

```
[22800 rows x 1 columns]
```

Adding 'death_reason' field

```
[827]: df_remv['death_reason'] = df_remv['cause'].str.slice(15,)
```

```
[828]: df_remv[['death_reason']]
```

```
[828]:
```

	death_reason
0	Gunfire
1	Gunfire
2	Gunfire
3	Gunfire
4	Gunfire
...	...
22795	Accidental
22796	Heat exhaustion
22797	Heat exhaustion
22798	Gunfire (Accidental)
22799	Automobile accident

```
[22800 rows x 1 columns]
```

```
[829]: df_remv
```

```
[829]:
```

	person \
0	Constable Darius Quimby
1	Sheriff Cornelius Hogeboom
2	Deputy Sheriff Isaac Smith
3	Marshal Robert Forsyth
4	Sheriff Robert Maxwell
...	...
22795	K9 Bruno
22796	K9 Lazer
22797	K9 Tyson
22798	K9 Credo
22799	K9 Roscoe

	dept \
0	Albany County Constable's Office, NY
1	Columbia County Sheriff's Office, NY
2	Westchester County Sheriff's Department, NY
3	United States Department of Justice - United S...


```

4           Greenville County Sheriff's Office, SC
...
22795           Amarillo Police Department, TX
22796 United States Department of Homeland Security ...
22797           Fountain County Sheriff's Office, IN
22798           Long Beach Police Department, CA
22799           Emmett Police Department, ID

```

```

                                eow                                cause \
0      EOW: Monday, January 3, 1791      Cause of Death: Gunfire
1      EOW: Saturday, October 22, 1791    Cause of Death: Gunfire
2      EOW: Thursday, May 17, 1792        Cause of Death: Gunfire
3      EOW: Saturday, January 11, 1794    Cause of Death: Gunfire
4      EOW: Sunday, November 12, 1797     Cause of Death: Gunfire
...
22795      EOW: Sunday, June 12, 2016      Cause of Death: Accidental
22796      EOW: Monday, June 20, 2016      Cause of Death: Heat exhaustion
22797      EOW: Monday, June 27, 2016      Cause of Death: Heat exhaustion
22798      EOW: Tuesday, June 28, 2016     Cause of Death: Gunfire (Accidental)
22799      EOW: Friday, July 1, 2016      Cause of Death: Automobile accident

```

```

                                year      date      month      death_reason
0      1791  01-03-1791  January      Gunfire
1      1791  10-22-1791  October      Gunfire
2      1792  05-17-1792   May          Gunfire
3      1794  01-11-1794  January      Gunfire
4      1797  11-12-1797  November     Gunfire
...
22795  2016  06-12-2016   June          Accidental
22796  2016  06-20-2016   June          Heat exhaustion
22797  2016  06-27-2016   June          Heat exhaustion
22798  2016  06-28-2016   June          Gunfire (Accidental)
22799  2016  07-01-2016   July           Automobile accident

```

[22800 rows x 8 columns]

Adding 'dept_name' & 'state' field

```

[863]: df_remv['dept_name']= df_remv['dept'].str.split(',').str[0]
df_remv['dept_name']=df_remv['dept_name'].str.strip()
#df_remv['state']= df_remv['dept'].str.split(',').str[1]
#df_remv['state']=df_remv['state'].str.strip()
df_remv['state']=df_remv['dept'].str.slice(-2,)

```

```

[864]: df_remv[['dept_name', 'state']]

```

```
[864]:
```

	dept_name	state
0	Albany County Constable's Office	NY
1	Columbia County Sheriff's Office	NY
2	Westchester County Sheriff's Department	NY
3	United States Department of Justice - United S...	US
4	Greenville County Sheriff's Office	SC
...
22795	Amarillo Police Department	TX
22796	United States Department of Homeland Security ...	US
22797	Fountain County Sheriff's Office	IN
22798	Long Beach Police Department	CA
22799	Emmett Police Department	ID

[22800 rows x 2 columns]

```
[865]: df_remv
```

```
[865]:
```

	person \
0	Constable Darius Quimby
1	Sheriff Cornelius Hogeboom
2	Deputy Sheriff Isaac Smith
3	Marshal Robert Forsyth
4	Sheriff Robert Maxwell
...	...
22795	K9 Bruno
22796	K9 Lazer
22797	K9 Tyson
22798	K9 Credo
22799	K9 Roscoe

	dept \
0	Albany County Constable's Office, NY
1	Columbia County Sheriff's Office, NY
2	Westchester County Sheriff's Department, NY
3	United States Department of Justice - United S...
4	Greenville County Sheriff's Office, SC
...	...
22795	Amarillo Police Department, TX
22796	United States Department of Homeland Security ...
22797	Fountain County Sheriff's Office, IN
22798	Long Beach Police Department, CA
22799	Emmett Police Department, ID

	eow	cause \
0	EOW: Monday, January 3, 1791	Cause of Death: Gunfire
1	EOW: Saturday, October 22, 1791	Cause of Death: Gunfire
2	EOW: Thursday, May 17, 1792	Cause of Death: Gunfire

3	EOW: Saturday, January 11, 1794	Cause of Death: Gunfire
4	EOW: Sunday, November 12, 1797	Cause of Death: Gunfire
...
22795	EOW: Sunday, June 12, 2016	Cause of Death: Accidental
22796	EOW: Monday, June 20, 2016	Cause of Death: Heat exhaustion
22797	EOW: Monday, June 27, 2016	Cause of Death: Heat exhaustion
22798	EOW: Tuesday, June 28, 2016	Cause of Death: Gunfire (Accidental)
22799	EOW: Friday, July 1, 2016	Cause of Death: Automobile accident

		dept_name	state
0		Albany County Constable's Office	NY
1		Columbia County Sheriff's Office	NY
2		Westchester County Sheriff's Department	NY
3	United States Department of Justice - United S...		US
4		Greenville County Sheriff's Office	SC
...	
22795		Amarillo Police Department	TX
22796	United States Department of Homeland Security ...		US
22797		Fountain County Sheriff's Office	IN
22798		Long Beach Police Department	CA
22799		Emmett Police Department	ID

[22800 rows x 6 columns]

```
[833]: df_data=df_remv[['person','dept','eow','cause','death_reason','date','month','year','dept_name
```

```
[834]: df=df_data
```

```
[866]: df
```

```
[866]:
           person \
0      Constable Darius Quimby
1    Sheriff Cornelius Hogeboom
2  Deputy Sheriff Isaac Smith
3      Marshal Robert Forsyth
4      Sheriff Robert Maxwell
...
22795          K9 Bruno
22796          K9 Lazer
22797          K9 Tyson
22798          K9 Credo
22799          K9 Roscoe
```

```

           dept \
0      Albany County Constable's Office, NY
1      Columbia County Sheriff's Office, NY
2  Westchester County Sheriff's Department, NY
```

3 United States Department of Justice - United S...
 4 Greenville County Sheriff's Office, SC
 ...
 22795 Amarillo Police Department, TX
 22796 United States Department of Homeland Security ...
 22797 Fountain County Sheriff's Office, IN
 22798 Long Beach Police Department, CA
 22799 Emmett Police Department, ID

	eow	cause \
0	EOW: Monday, January 3, 1791	Cause of Death: Gunfire
1	EOW: Saturday, October 22, 1791	Cause of Death: Gunfire
2	EOW: Thursday, May 17, 1792	Cause of Death: Gunfire
3	EOW: Saturday, January 11, 1794	Cause of Death: Gunfire
4	EOW: Sunday, November 12, 1797	Cause of Death: Gunfire
...
22795	EOW: Sunday, June 12, 2016	Cause of Death: Accidental
22796	EOW: Monday, June 20, 2016	Cause of Death: Heat exhaustion
22797	EOW: Monday, June 27, 2016	Cause of Death: Heat exhaustion
22798	EOW: Tuesday, June 28, 2016	Cause of Death: Gunfire (Accidental)
22799	EOW: Friday, July 1, 2016	Cause of Death: Automobile accident

	death_reason	date	month	year \
0	Gunfire	01-03-1791	January	1791
1	Gunfire	10-22-1791	October	1791
2	Gunfire	05-17-1792	May	1792
3	Gunfire	01-11-1794	January	1794
4	Gunfire	11-12-1797	November	1797
...
22795	Accidental	06-12-2016	June	2016
22796	Heat exhaustion	06-20-2016	June	2016
22797	Heat exhaustion	06-27-2016	June	2016
22798	Gunfire (Accidental)	06-28-2016	June	2016
22799	Automobile accident	07-01-2016	July	2016

	dept_name	state
0	Albany County Constable's Office	NY
1	Columbia County Sheriff's Office	NY
2	Westchester County Sheriff's Department	NY
3	United States Department of Justice - United S...	US
4	Greenville County Sheriff's Office	SC
...
22795	Amarillo Police Department	TX
22796	United States Department of Homeland Security ...	US
22797	Fountain County Sheriff's Office	IN
22798	Long Beach Police Department	CA
22799	Emmett Police Department	ID

[22800 rows x 10 columns]

```
[836]: newdf = df[(df.year >= 1900) & (df.year <= 2021)]
```

```
[837]: newdf
```

```
[837]:
```

	person \		dept \		eow	cause \
1917	Constable W. D. Turner		Lauderdale County Constable's Office, TN		EOW: Tuesday, January 9, 1900	Cause of Death: Gunfire
1918	Deputy Constable Marvin Durham		Lauderdale County Constable's Office, TN		EOW: Tuesday, January 9, 1900	Cause of Death: Gunfire
1919	Jailer Alfred Henry		Howell County Sheriff's Department, MO		EOW: Wednesday, January 17, 1900	Cause of Death: Assault
1920	Night Captain William C. Rooney		Colorado Department of Corrections, CO		EOW: Monday, January 22, 1900	Cause of Death: Stabbed
1921	Deputy Constable George W. McCammon		Pennsylvania State Constable - Washington Coun...		EOW: Monday, January 22, 1900	Cause of Death: Gunfire
...
22795	K9 Bruno		Amarillo Police Department, TX		EOW: Sunday, June 12, 2016	Cause of Death: Accidental
22796	K9 Lazer		United States Department of Homeland Security ...		EOW: Monday, June 20, 2016	Cause of Death: Heat exhaustion
22797	K9 Tyson		Fountain County Sheriff's Office, IN		EOW: Monday, June 27, 2016	Cause of Death: Heat exhaustion
22798	K9 Credo		Long Beach Police Department, CA		EOW: Tuesday, June 28, 2016	Cause of Death: Gunfire (Accidental)
22799	K9 Roscoe		Emmett Police Department, ID		EOW: Friday, July 1, 2016	Cause of Death: Automobile accident
					death_reason	date month year \

1917	Gunfire	01-09-1900	January	1900
1918	Gunfire	01-09-1900	January	1900
1919	Assault	01-17-1900	January	1900
1920	Stabbed	01-22-1900	January	1900
1921	Gunfire	01-22-1900	January	1900
...
22795	Accidental	06-12-2016	June	2016
22796	Heat exhaustion	06-20-2016	June	2016
22797	Heat exhaustion	06-27-2016	June	2016
22798	Gunfire (Accidental)	06-28-2016	June	2016
22799	Automobile accident	07-01-2016	July	2016

			dept_name	state
1917		Lauderdale County Constable's Office		TN
1918		Lauderdale County Constable's Office		TN
1919		Howell County Sheriff's Department		MO
1920		Colorado Department of Corrections		CO
1921	Pennsylvania State Constable - Washington County			PA
...		
22795		Amarillo Police Department		TX
22796	United States Department of Homeland Security ...			US
22797		Fountain County Sheriff's Office		IN
22798		Long Beach Police Department		CA
22799		Emmett Police Department		ID

[20883 rows x 10 columns]

3. Analysis Using Matplotlib library

```
[14]: !pip install seaborn
```

```
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution - (/opt/conda/lib/python3.7/site-
packages)
WARNING: Ignoring invalid distribution -ertifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution - (/opt/conda/lib/python3.7/site-
packages)
WARNING: Ignoring invalid distribution -ertifi
(/opt/conda/lib/python3.7/site-packages)
Requirement already satisfied: seaborn in /opt/conda/lib/python3.7/site-packages
```

```

(0.11.2)
Requirement already satisfied: numpy>=1.15 in /opt/conda/lib/python3.7/site-
packages (from seaborn) (1.19.2)
Requirement already satisfied: scipy>=1.0 in /opt/conda/lib/python3.7/site-
packages (from seaborn) (1.5.2)
Requirement already satisfied: matplotlib>=2.2 in /opt/conda/lib/python3.7/site-
packages (from seaborn) (3.3.2)
Requirement already satisfied: pandas>=0.23 in /opt/conda/lib/python3.7/site-
packages (from seaborn) (1.1.2)
Requirement already satisfied: python-dateutil>=2.1 in
/opt/conda/lib/python3.7/site-packages (from matplotlib>=2.2->seaborn) (2.8.1)
Collecting certifi>=2020.06.20
  Using cached certifi-2021.10.8-py2.py3-none-any.whl (149 kB)
Requirement already satisfied: cycler>=0.10 in /opt/conda/lib/python3.7/site-
packages (from matplotlib>=2.2->seaborn) (0.10.0)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.3 in
/opt/conda/lib/python3.7/site-packages (from matplotlib>=2.2->seaborn) (2.4.7)
Requirement already satisfied: kiwisolver>=1.0.1 in
/opt/conda/lib/python3.7/site-packages (from matplotlib>=2.2->seaborn) (1.2.0)
Requirement already satisfied: pillow>=6.2.0 in /opt/conda/lib/python3.7/site-
packages (from matplotlib>=2.2->seaborn) (7.2.0)
Requirement already satisfied: pytz>=2017.2 in /opt/conda/lib/python3.7/site-
packages (from pandas>=0.23->seaborn) (2020.1)
Requirement already satisfied: six in /opt/conda/lib/python3.7/site-packages
(from cycler>=0.10->matplotlib>=2.2->seaborn) (1.14.0)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution - (/opt/conda/lib/python3.7/site-
packages)
WARNING: Ignoring invalid distribution -ertifi
(/opt/conda/lib/python3.7/site-packages)
Installing collected packages: certifi
  Attempting uninstall: certifi
    WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
    WARNING: Ignoring invalid distribution -
(/opt/conda/lib/python3.7/site-packages)
    WARNING: Ignoring invalid distribution -ertifi
(/opt/conda/lib/python3.7/site-packages)
    Found existing installation: certifi 2019.11.28
    Uninstalling certifi-2019.11.28:

```

```
ERROR: Could not install packages due to an OSError: [Errno 13] Permission
denied: 'metadata.json'
```

```
Consider using the `--user` option or check the permissions.
```

```
WARNING: Ignoring invalid distribution -rtifi
```

```
(/opt/conda/lib/python3.7/site-packages)
```

```
WARNING: Ignoring invalid distribution - (/opt/conda/lib/python3.7/site-
packages)
```

```
WARNING: Ignoring invalid distribution -ertifi
```

```
(/opt/conda/lib/python3.7/site-packages)
```

```
WARNING: Ignoring invalid distribution -rtifi
```

```
(/opt/conda/lib/python3.7/site-packages)
```

```
WARNING: Ignoring invalid distribution - (/opt/conda/lib/python3.7/site-
packages)
```

```
WARNING: Ignoring invalid distribution -ertifi
```

```
(/opt/conda/lib/python3.7/site-packages)
```

```
[838]: import matplotlib.pyplot as plt
import seaborn as sns
```

```
[839]: newdf = df[(df.year >= 1900) & (df.year <= 2021)]
```

```
[840]: newdf
```

```
[840]:
```

	person \	dept \
1917	Constable W. D. Turner	
1918	Deputy Constable Marvin Durham	
1919	Jailer Alfred Henry	
1920	Night Captain William C. Rooney	
1921	Deputy Constable George W. McCammon	
...	...	
22795	K9 Bruno	
22796	K9 Lazer	
22797	K9 Tyson	
22798	K9 Credo	
22799	K9 Roscoe	
1917		Lauderdale County Constable's Office, TN
1918		Lauderdale County Constable's Office, TN
1919		Howell County Sheriff's Department, MO
1920		Colorado Department of Corrections, CO

1921 Pennsylvania State Constable - Washington Coun...

...
22795 Amarillo Police Department, TX
22796 United States Department of Homeland Security ...
22797 Fountain County Sheriff's Office, IN
22798 Long Beach Police Department, CA
22799 Emmett Police Department, ID

	eow	cause \
1917	EOW: Tuesday, January 9, 1900	Cause of Death: Gunfire
1918	EOW: Tuesday, January 9, 1900	Cause of Death: Gunfire
1919	EOW: Wednesday, January 17, 1900	Cause of Death: Assault
1920	EOW: Monday, January 22, 1900	Cause of Death: Stabbed
1921	EOW: Monday, January 22, 1900	Cause of Death: Gunfire
...
22795	EOW: Sunday, June 12, 2016	Cause of Death: Accidental
22796	EOW: Monday, June 20, 2016	Cause of Death: Heat exhaustion
22797	EOW: Monday, June 27, 2016	Cause of Death: Heat exhaustion
22798	EOW: Tuesday, June 28, 2016	Cause of Death: Gunfire (Accidental)
22799	EOW: Friday, July 1, 2016	Cause of Death: Automobile accident

	death_reason	date	month	year \
1917	Gunfire	01-09-1900	January	1900
1918	Gunfire	01-09-1900	January	1900
1919	Assault	01-17-1900	January	1900
1920	Stabbed	01-22-1900	January	1900
1921	Gunfire	01-22-1900	January	1900
...
22795	Accidental	06-12-2016	June	2016
22796	Heat exhaustion	06-20-2016	June	2016
22797	Heat exhaustion	06-27-2016	June	2016
22798	Gunfire (Accidental)	06-28-2016	June	2016
22799	Automobile accident	07-01-2016	July	2016

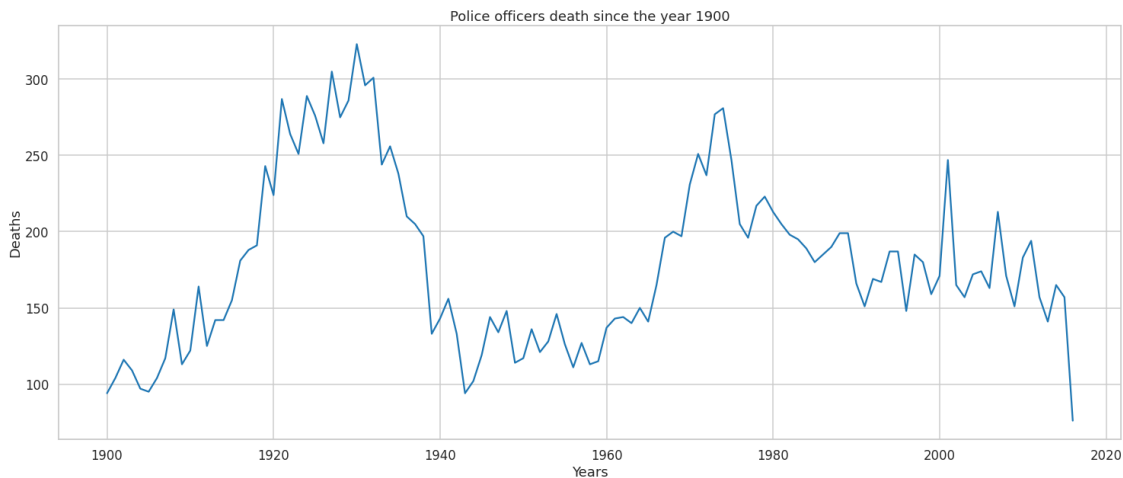
	dept_name	state
1917	Lauderdale County Constable's Office	TN
1918	Lauderdale County Constable's Office	TN
1919	Howell County Sheriff's Department	MO
1920	Colorado Department of Corrections	CO
1921	Pennsylvania State Constable - Washington County	PA
...
22795	Amarillo Police Department	TX
22796	United States Department of Homeland Security ...	US
22797	Fountain County Sheriff's Office	IN
22798	Long Beach Police Department	CA
22799	Emmett Police Department	ID

[20883 rows x 10 columns]

```
[867]: yeardf=newdf.year.value_counts().sort_index()
plt.figure(figsize=(25,10))
sns.lineplot(yeardf.index, yeardf.values)
plt.title('Police officers death since the year 1900')
plt.xlabel('Years')
plt.ylabel('Deaths')
plt.show()
```

/opt/conda/lib/python3.7/site-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

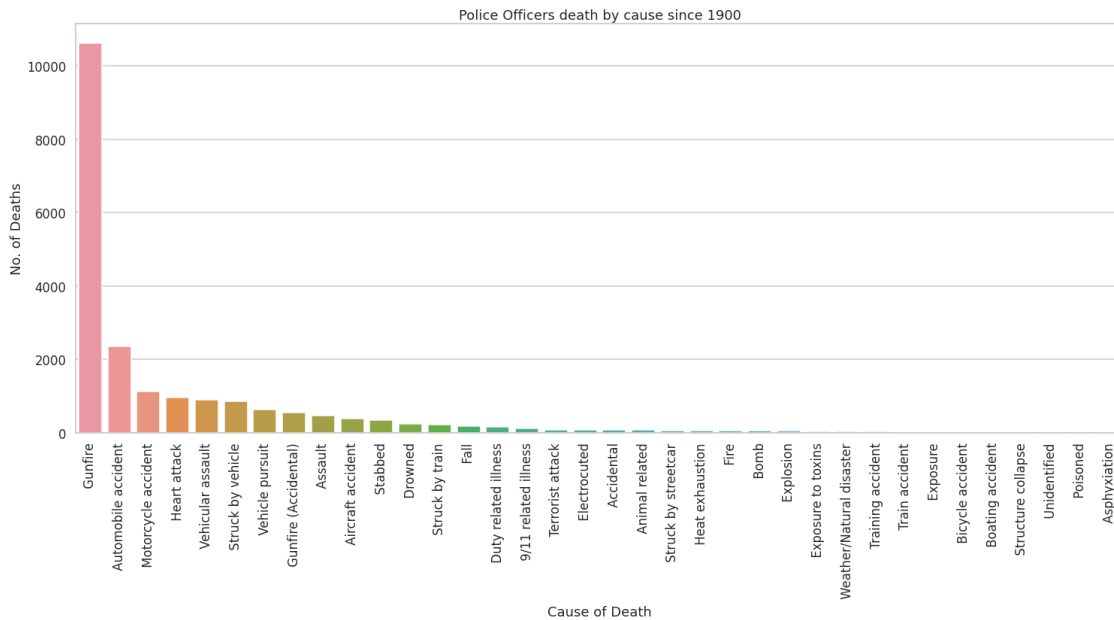
FutureWarning



```
[868]: causeddf=newdf.death_reason.value_counts()
plt.figure(figsize=(25,10))
sns.barplot(causeddf.index, causeddf.values)
plt.title('Police Officers death by cause since 1900')
plt.xlabel('Cause of Death')
plt.ylabel('No. of Deaths')
plt.xticks(rotation=90)
plt.show()
```

/opt/conda/lib/python3.7/site-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

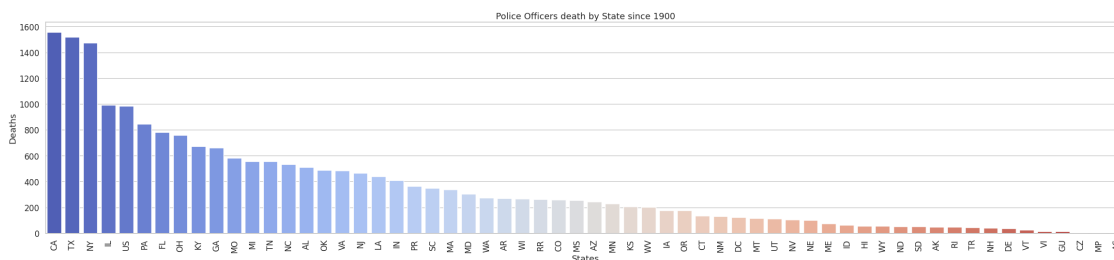
FutureWarning



```
[869]: stateddf=newdf.state.value_counts()
plt.figure(figsize=(40,8))
sns.barplot(stateddf.index,stateddf.values,palette='coolwarm')
plt.title('Police Officers death by State since 1900')
plt.xlabel('States')
plt.ylabel('Deaths')
plt.xticks(rotation=90)
plt.show()
```

/opt/conda/lib/python3.7/site-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

FutureWarning

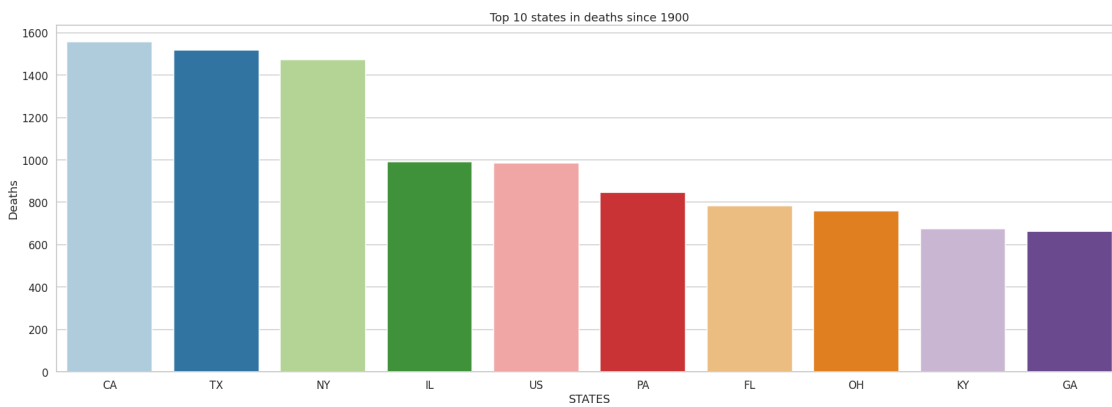


```
[870]: cntdf=newdf['state'].value_counts()
topdf=cntdf.head(10)
```

```
plt.figure(figsize=(30,10))
sns.barplot(topdf.index,topdf.values,palette='Paired')
plt.title('Top 10 states in deaths since 1900')
plt.xlabel('STATES')
plt.ylabel('Deaths')
plt.show()
```

/opt/conda/lib/python3.7/site-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

FutureWarning



4. ML Modeling

[234]: `pip install scikit-learn`

```
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution - (/opt/conda/lib/python3.7/site-
packages)
WARNING: Ignoring invalid distribution -ertifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -cipy
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
```

```

WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution - (/opt/conda/lib/python3.7/site-
packages)
WARNING: Ignoring invalid distribution -ertifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -cipy
(/opt/conda/lib/python3.7/site-packages)
Collecting scikit-learn
  Using cached
scikit_learn-1.0.1-cp37-cp37m-manylinux_2_12_x86_64.manylinux2010_x86_64.whl
(23.2 MB)
Requirement already satisfied: numpy>=1.14.6 in /opt/conda/lib/python3.7/site-
packages (from scikit-learn) (1.19.2)
Requirement already satisfied: scipy>=1.1.0 in /opt/conda/lib/python3.7/site-
packages (from scikit-learn) (1.5.2)
Collecting joblib>=0.11
  Using cached joblib-1.1.0-py2.py3-none-any.whl (306 kB)
Requirement already satisfied: threadpoolctl>=2.0.0 in
/opt/conda/lib/python3.7/site-packages (from scikit-learn) (3.0.0)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution - (/opt/conda/lib/python3.7/site-
packages)
WARNING: Ignoring invalid distribution -ertifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -cipy
(/opt/conda/lib/python3.7/site-packages)
Installing collected packages: joblib, scikit-learn
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution - (/opt/conda/lib/python3.7/site-
packages)
WARNING: Ignoring invalid distribution -ertifi
(/opt/conda/lib/python3.7/site-packages)

```

```

WARNING: Ignoring invalid distribution -cipy
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution - (/opt/conda/lib/python3.7/site-
packages)
WARNING: Ignoring invalid distribution -ertifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -cipy
(/opt/conda/lib/python3.7/site-packages)
Successfully installed joblib-1.1.0 scikit-learn-1.0.1
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution - (/opt/conda/lib/python3.7/site-
packages)
WARNING: Ignoring invalid distribution -ertifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -cipy
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution - (/opt/conda/lib/python3.7/site-
packages)
WARNING: Ignoring invalid distribution -ertifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -cipy
(/opt/conda/lib/python3.7/site-packages)
Note: you may need to restart the kernel to use updated packages.

```

1. Implementing Decision Tree Classifier

1 The feature input labels are 'year' and 'deptcode'. The output label is 'death_reason'

```
[846]: newdf['statecode']=newdf.state.astype("category").cat.codes
newdf['deptcode']=newdf.dept_name.astype("category").cat.codes
newdf['deathcode']=newdf.death_reason.astype("category").cat.codes
```

```
/opt/conda/lib/python3.7/site-packages/ipykernel_launcher.py:1:
```

```
SettingWithCopyWarning:
```

```
A value is trying to be set on a copy of a slice from a DataFrame.
```

```
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
"""Entry point for launching an IPython kernel.
```

```
/opt/conda/lib/python3.7/site-packages/ipykernel_launcher.py:2:
```

```
SettingWithCopyWarning:
```

```
A value is trying to be set on a copy of a slice from a DataFrame.
```

```
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
/opt/conda/lib/python3.7/site-packages/ipykernel_launcher.py:3:
```

```
SettingWithCopyWarning:
```

```
A value is trying to be set on a copy of a slice from a DataFrame.
```

```
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

This is separate from the ipykernel package so we can avoid doing imports until

```
[847]: newdf[['dept', 'deptcode']]
```

```
[847]:
```

	dept	deptcode
1917	Lauderdale County Constable's Office, TN	2460
1918	Lauderdale County Constable's Office, TN	2460
1919	Howell County Sheriff's Department, MO	2084
1920	Colorado Department of Corrections, CO	926
1921	Pennsylvania State Constable - Washington Coun...	3622
...
22795	Amarillo Police Department, TX	97
22796	United States Department of Homeland Security ...	4702
22797	Fountain County Sheriff's Office, IN	1570
22798	Long Beach Police Department, CA	2585

22799

Emmett Police Department, ID

1380

[20883 rows x 2 columns]

```
[848]: #X=new2df.  
        ↳drop(columns=['person','dept','eow','cause','date','month','dept_name','state','dcode'])  
        #X=new2df.  
        ↳drop(columns=['person','dept','eow','cause','date','month','dept_name','death_reason','dcode'])  
        X=newdf.  
        ↳drop(columns=['person','dept','eow','cause','date','month','dept_name','death_reason','death_reason'])  
        X
```

```
[848]:      year  deptcode  
1917   1900      2460  
1918   1900      2460  
1919   1900      2084  
1920   1900       926  
1921   1900     3622  
...    ...    ...  
22795  2016        97  
22796  2016     4702  
22797  2016     1570  
22798  2016     2585  
22799  2016     1380
```

[20883 rows x 2 columns]

```
[849]: import sklearn  
        from sklearn.tree import DecisionTreeClassifier  
        from sklearn.model_selection import train_test_split  
        #from sklearn.metrics import accuracy_score  
  
        #X=new2df.  
        ↳drop(columns=['person','dept','eow','cause','date','month','dept_name','state','death_reason'])  
        y=newdf['death_reason']  
        X_train, X_test, y_train, y_test =train_test_split(X,y,test_size=0.2)  
  
        model=DecisionTreeClassifier()  
        model.fit(X_train, y_train)  
        predictions=model.predict(X_test)
```

```
[809]: from sklearn.metrics import accuracy_score  
        accuracy_score(y_test,predictions)*100
```

75.81134785731386


```
[850]: #X=new2df.
        ↳drop(columns=['person','dept','eow','cause','date','month','dept_name','state','dcode'])
        #X=new2df.
        ↳drop(columns=['person','dept','eow','cause','date','month','dept_name','death_reason','dcode'])
        X=newdf.
        ↳drop(columns=['person','dept','eow','cause','date','month','dept_name','death_reason','death_reason'])
        X
```

```
[850]:      year  statecode
1917   1900         48
1918   1900         48
1919   1900         27
1920   1900          6
1921   1900         42
...
22795  2016         50
22796  2016         51
22797  2016         18
22798  2016          5
22799  2016         16

[20883 rows x 2 columns]
```

2 The feature input labels are ‘year’ and ‘statecode’. The output label is ‘death_reason’

```
[851]: newdf[['state','statecode']]
```

```
[851]:      state  statecode
1917     TN         48
1918     TN         48
1919     MO         27
1920     CO          6
1921     PA         42
...
22795     TX         50
22796     US         51
22797     IN         18
22798     CA          5
22799     ID         16

[20883 rows x 2 columns]
```

```
[852]: import sklearn
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
#from sklearn.metrics import accuracy_score

#X=new2df.
↳drop(columns=['person','dept','eow','cause','date','month','dept_name','state','death_reaso
y=newdf['death_reason']
X_train, X_test, y_train, y_test =train_test_split(X,y,test_size=0.2)

model=DecisionTreeClassifier()
model.fit(X_train, y_train)
predictions=model.predict(X_test)
```

```
[810]: from sklearn.metrics import accuracy_score
accuracy_score(y_test,predictions)
```

81.24869400526694

2. Implementing K-Nearest Neighbours Classifier

3 The feature input labels are ‘year’ and ‘statecode’. The output label is ‘death_reason’

```
[871]: #X=new2df.
↳drop(columns=['person','dept','eow','cause','date','month','dept_name','state','dcode'])
#X=new2df.
↳drop(columns=['person','dept','eow','cause','date','month','dept_name','death_reason','dcode'])
X=newdf.
↳drop(columns=['person','dept','eow','cause','date','month','dept_name','death_reason','death_reason'])
```

```
[877]: from sklearn import linear_model
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier(n_neighbors=15)
```

```
[878]: knn.fit(X_train, y_train)
```

```
[878]: KNeighborsClassifier(n_neighbors=15)
```

```
[879]: knnpred=knn.predict(X_test)
```

```
[880]: accuracy_score(y_test,knpred)*100
```

```
[880]: 50.58654536748863
```

[495]: `pip install requests`

```
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution - (/opt/conda/lib/python3.7/site-
packages)
WARNING: Ignoring invalid distribution -ertifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -cipy
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution - (/opt/conda/lib/python3.7/site-
packages)
WARNING: Ignoring invalid distribution -ertifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -cipy
(/opt/conda/lib/python3.7/site-packages)
Requirement already satisfied: requests in /opt/conda/lib/python3.7/site-
packages (2.22.0)
Requirement already satisfied: urllib3!=1.25.0,!1.25.1,<1.26,>=1.21.1 in
/opt/conda/lib/python3.7/site-packages (from requests) (1.25.7)
Requirement already satisfied: idna<2.9,>=2.5 in /opt/conda/lib/python3.7/site-
packages (from requests) (2.8)
Requirement already satisfied: chardet<3.1.0,>=3.0.2 in
/opt/conda/lib/python3.7/site-packages (from requests) (3.0.4)
Requirement already satisfied: certifi>=2017.4.17 in
/opt/conda/lib/python3.7/site-packages (from requests) (2019.11.28)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution - (/opt/conda/lib/python3.7/site-
packages)
WARNING: Ignoring invalid distribution -ertifi
(/opt/conda/lib/python3.7/site-packages)
```

```

WARNING: Ignoring invalid distribution -cipy
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution - (/opt/conda/lib/python3.7/site-
packages)
WARNING: Ignoring invalid distribution -ertifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -cipy
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution - (/opt/conda/lib/python3.7/site-
packages)
WARNING: Ignoring invalid distribution -ertifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -cipy
(/opt/conda/lib/python3.7/site-packages)
Note: you may need to restart the kernel to use updated packages.

```

[496]: `pip install bs4`

```

WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution - (/opt/conda/lib/python3.7/site-
packages)
WARNING: Ignoring invalid distribution -ertifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -cipy
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)

```

```

WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution - (/opt/conda/lib/python3.7/site-
packages)
WARNING: Ignoring invalid distribution -ertifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -cipy
(/opt/conda/lib/python3.7/site-packages)
Collecting bs4
  Downloading bs4-0.0.1.tar.gz (1.1 kB)
  Preparing metadata (setup.py) ... done
Collecting beautifulsoup4
  Downloading beautifulsoup4-4.10.0-py3-none-any.whl (97 kB)
    |                               | 97 kB 412 kB/s
Collecting soupsieve>1.2
  Downloading soupsieve-2.3.1-py3-none-any.whl (37 kB)
Building wheels for collected packages: bs4
  Building wheel for bs4 (setup.py) ... done
  Created wheel for bs4: filename=bs4-0.0.1-py3-none-any.whl size=1272
sha256=13f76681bb6866678dee8eef1fb0ba809c8e79a11f5992a323ef07f4dc68a719
  Stored in directory: /home/jovyan/.cache/pip/wheels/0a/9e/ba/20e5bbc1afef3a491
f0b3bb74d508f99403aabe76eda2167ca
Successfully built bs4
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution - (/opt/conda/lib/python3.7/site-
packages)
WARNING: Ignoring invalid distribution -ertifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -cipy
(/opt/conda/lib/python3.7/site-packages)
Installing collected packages: soupsieve, beautifulsoup4, bs4
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution - (/opt/conda/lib/python3.7/site-
packages)

```

```

WARNING: Ignoring invalid distribution -ertifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -cipy
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution - (/opt/conda/lib/python3.7/site-
packages)
WARNING: Ignoring invalid distribution -ertifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -cipy
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution - (/opt/conda/lib/python3.7/site-
packages)
WARNING: Ignoring invalid distribution -ertifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -cipy
(/opt/conda/lib/python3.7/site-packages)
Successfully installed beautifulsoup4-4.10.0 bs4-0.0.1 soupsieve-2.3.1
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution - (/opt/conda/lib/python3.7/site-
packages)
WARNING: Ignoring invalid distribution -ertifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -cipy
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)

```

```
WARNING: Ignoring invalid distribution -rtifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution - (/opt/conda/lib/python3.7/site-
packages)
WARNING: Ignoring invalid distribution -ertifi
(/opt/conda/lib/python3.7/site-packages)
WARNING: Ignoring invalid distribution -cipy
(/opt/conda/lib/python3.7/site-packages)
Note: you may need to restart the kernel to use updated packages.
```