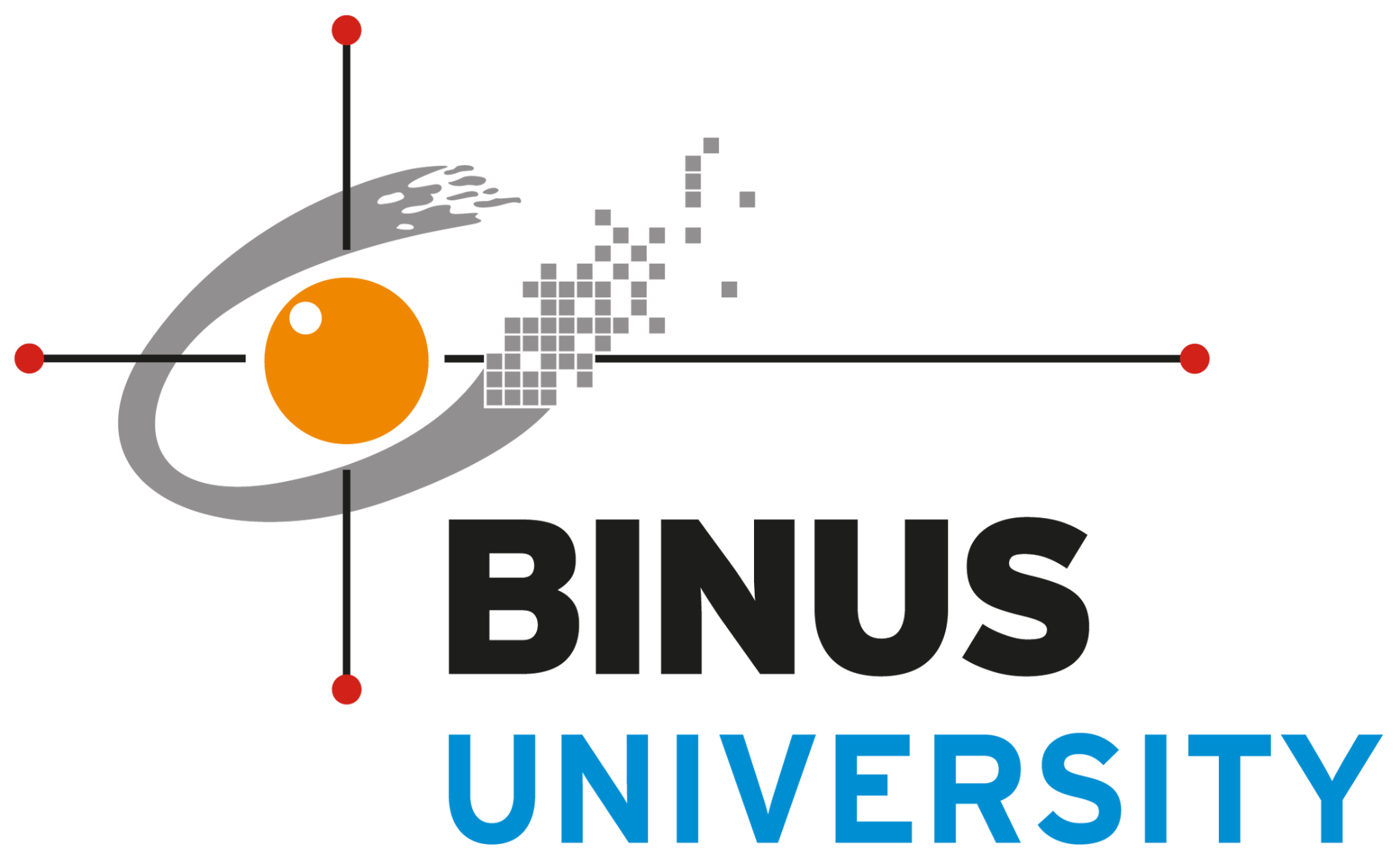
# **Mass Shootings in United States (2018-2022) Analytics**

A Project Report of

Big Data Processing Course



Submitted by:

2440047173 - Alexander Ryu Fenando

2440004395 - Even Owen Thamrin

2440003770 - Kent Samuel Claudio

2440030456 - Vetri Marvel Budiman

2440018141 - Yulius Obi

COMPUTER SCIENCE MAJOR

BINA NUSANTARA UNIVERSITY

2022

1. **Latar belakang masalah**

Penembakan massal adalah insiden di mana empat orang atau lebih (tidak termasuk si penembak) tertembak dan terbunuh. Jumlah penembakan massal yang melanda negara Amerika Serikat terlalu tinggi. Dari tahun 2021 saja sudah terdapat 692 jumlah kasus penembakan, yang menimbulkan 3519 korban yang terluka maupun korban jiwa.

1 dari 4 korban penembakan massal adalah anak-anak atau remaja. Adegan mengerikan penembakan massal telah menghantui masyarakat Amerika Serikat. Banyaknya regulasi mengenai kepemilikan senjata api di setiap negara bagian belum bisa menghapus insiden penembakan massal. Oleh karena itu, kami membuat analisis penembakan massal di setiap negara bagian Amerika Serikat untuk memberikan gambaran mengenai jumlah kasus penembakan massal yang ada di setiap negara bagian Amerika Serikat, serta meningkatkan kesadaran akan bahayanya kepemilikan senjata api dan dampak buruknya.

1. **Tujuan**

* Menampilkan jumlah kasus dan jumlah korban pada insiden penembakan massal di setiap negara bagian di Amerika Serikat dalam rentang 2018 - 2022.
* Mencari tahu dan menampilkan hari, bulan, tahun, musim dan public holiday yang paling sering terjadi insiden penembakan massal dalam rentang 2018 - 2022.
* Menampilkan dan menganalisa apa yang bisa didapat dari deskripsi terjadinya penembakan massal.
* Menampilkan informasi terkait keketatan regulasi mengenai penggunaan senjata api di Amerika Serikat berdasarkan banyaknya regulasi di setiap negara bagian pada tahun 2017.

1. **Manfaat**

* Meningkatkan kesadaran akan bahayanya kepemilikan senjata api dan dampak buruknya.
* Mengetahui ketat dan longgarnya aturan mengenai kepemilikan dan penggunaan senjata api di setiap negara bagian.

1. **Metodologi**
   1. Dataset 1

Kami menggunakan dataset yang berasal dari website Kaggle yang berjudul “Mass Shooting in United States (2018 - 2022).” Dataset ini mengandung dokumentasi kasus penembakan massal di setiap negara bagian Amerika Serikat dari tahun 2018 - 2022.

1. **Contain**

Dataset ini memiliki 5 file dengan format csv yang dibedakan dengan tahun dari daftar penembakan massal yaitu dari tahun 2018 sampai tahun 2022

1. **Columns**

Setiap file memiliki jumlah kolom yang sama yaitu 6 kolom. Kolom tersebut antara lain:

* 1. Date: berisi tanggal terjadinya penembakan massal
     1. Tipe data: date
  2. State: berisi dimana penembakkan massal terjadi
     1. Tipe data: string
  3. Dead: jumlah korban meninggal dari penembakkan massal
     1. Tipe data: integer
  4. Injured: jumlah korban terluka dari penembakkan massal
     1. Tipe data: integer
  5. Total: total dari jumlah korban terluka dan meninggal
     1. Tipe data: integer
  6. Description: deskripsi atau laporan singkat mengenai detail bagaimana terjadinya penembakan tersebut
     1. Tipe data: string

Sumber Dataset:<https://www.kaggle.com/datasets/hemil26/mass-shootings-in-united-states-20182022?select=shootings_2022.csv>

* 1. Dataset 2

Kami juga menggunakan dataset yang berasal dari website Kaggle yang berjudul “Firearms Provisions in US States”. Dataset ini mengandung jumlah regulasi tentang kepemilikan senjata api di setiap negara bagian Amerika Serikat.

1. **Contain**

Dataset ini memiliki 2 file dengan format csv dan xls yang dibagi menjadi negara yang memiliki regulasi bersenjata dari daftar regulasi yang berlaku dari tahun 1991 sampai tahun 2017. Kami menggunakan file raw\_data.csv yang tersedia.

1. **Columns**

Dalam file ‘raw\_data.csv’ memiliki jumlah kolom sebanyak 136 kolom. Namun, kami hanya memakai kolom yang kami perlukan yaitu ‘state’, ‘year’, dan ‘lawtotal’. Deskripsi dari kolom-kolom tersebut adalah sebagai berikut;

* 1. state: berisi dimana penembakkan massal terjadi
     1. Tipe data: String (nama negara bagian)
  2. year: berisi tahun regulasi tersebut dikemukakan.
     1. Tipe data: integer
  3. lawtotal: berisi jumlah regulasi yang berlaku mengenai senjata api
     1. Tipe data: number

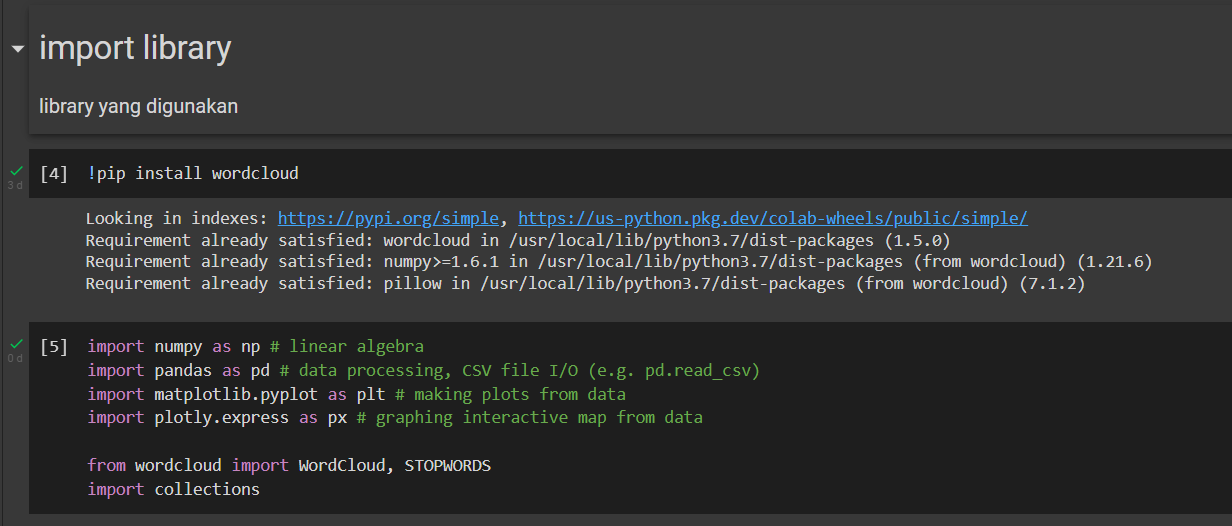
Sumber dataset:

<https://www.kaggle.com/datasets/jboysen/state-firearms>

* 1. Tools

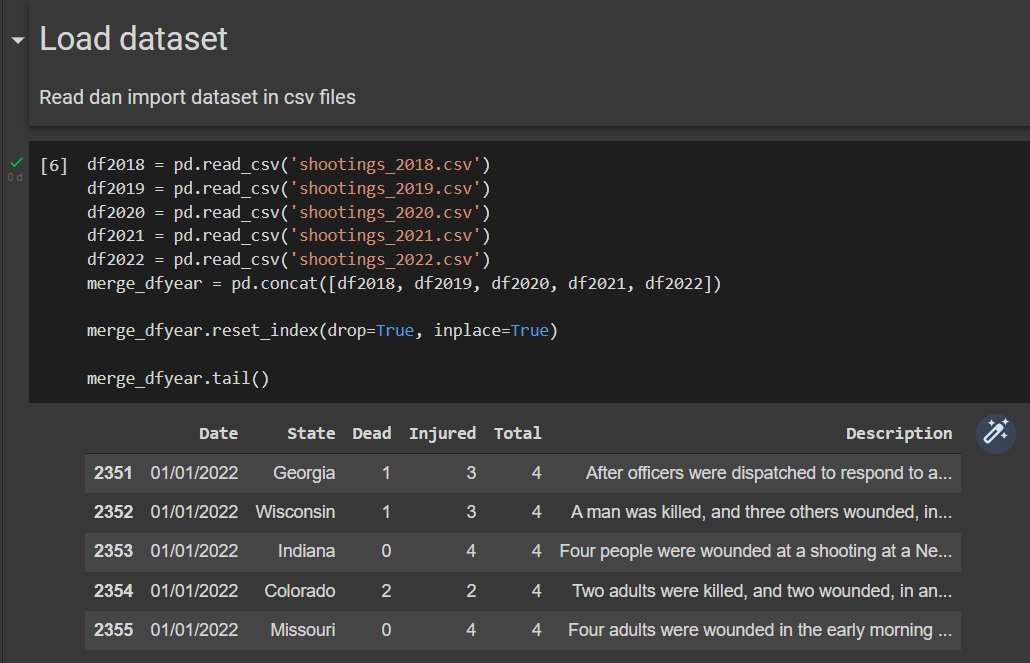
Kami menggunakan Google Colab dan bahasa pemrograman Python untuk melakukan analysis

1. **Processing data**
   1. Import library yang diperlukan
2. numpy
3. pandas
4. matplotlib.pyplot
5. ploty.express
6. wordcloud
7. collections



**Gambar 1**

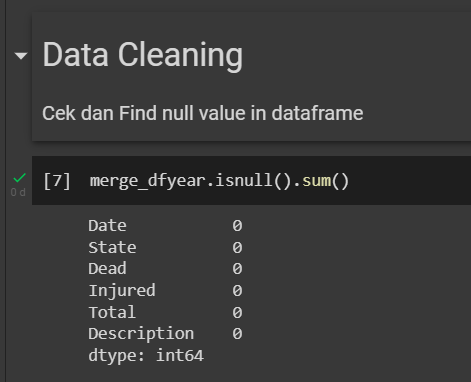
* 1. Melakukan download dataset dari Kaggle ke local komputer
  2. Memuat dan membaca dataset ke Google colab



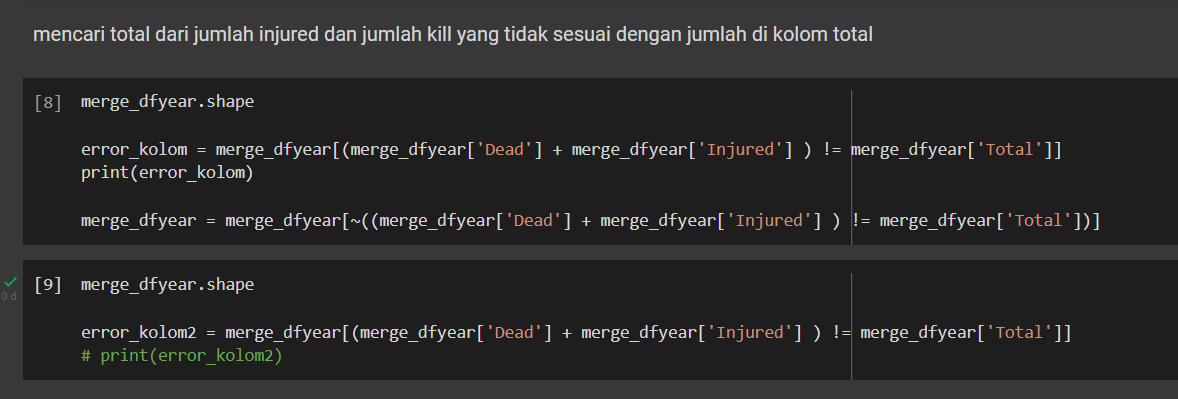
**Gambar 2**

* 1. Melakukan data cleaning

1. Melakukan reset index pada dataframe
2. Mengecek dan mencari apakah terdapat data yang memiliki value null
3. Mencari keakuratan dari kolom Total yang didapat dari jumlah kolom Dead dengan kolom Injured. Setelah menemukan data yang tidak sesuai atau tidak akurat, kami menghapus data tersebut.

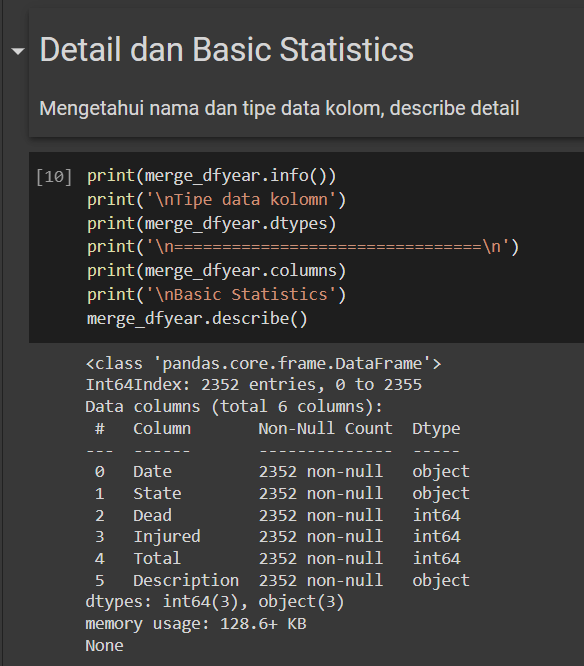


**Gambar 3**

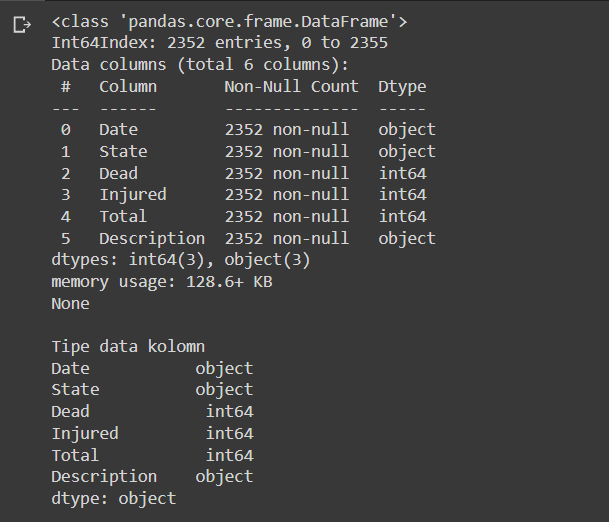


**Gambar 4**

* 1. Menampilkan tipe data dan informasi kolom dataset

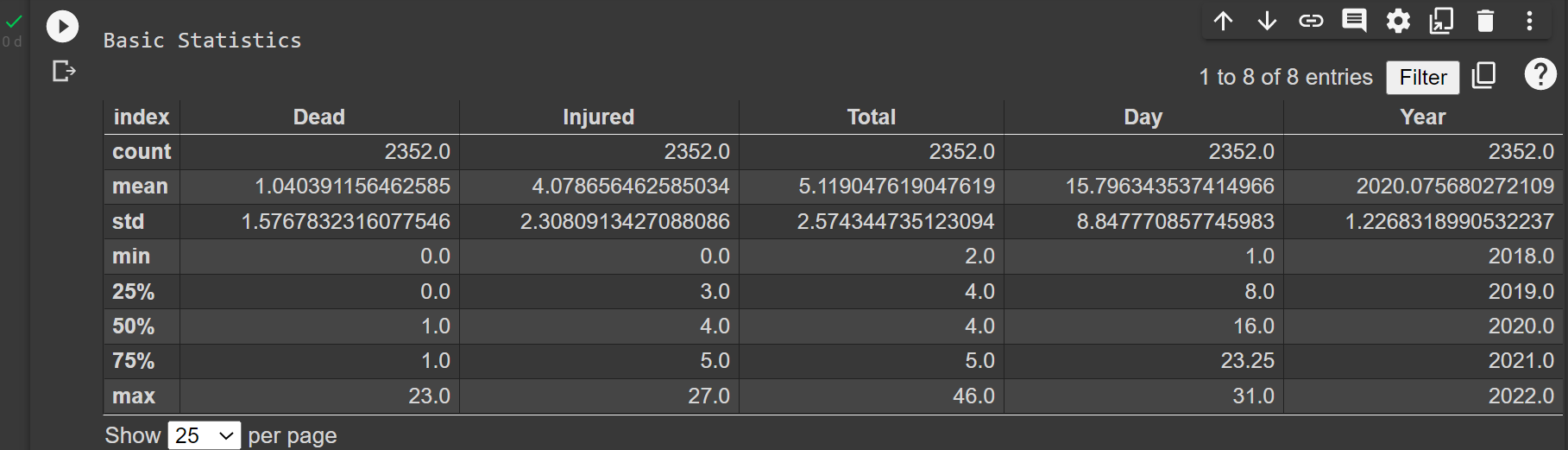


**Gambar 5**



**Gambar 6**

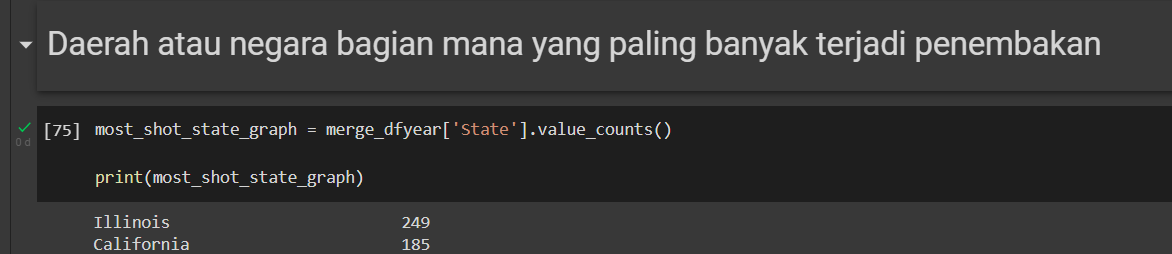
* 1. Menampilkan data basic statistics seperti count, mean, dan lainnya



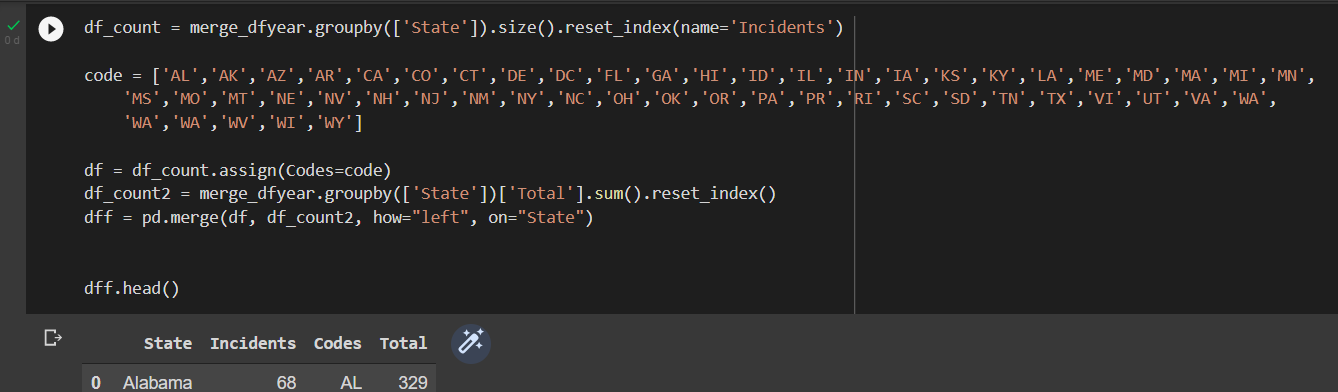
**Gambar 7**

* 1. Selanjutnya, kami melakukan analisis dengan 4 poin utama yaitu

1. Menganalisis jumlah kasus penembakan dan jumlah korban penembakan di setiap negara bagian



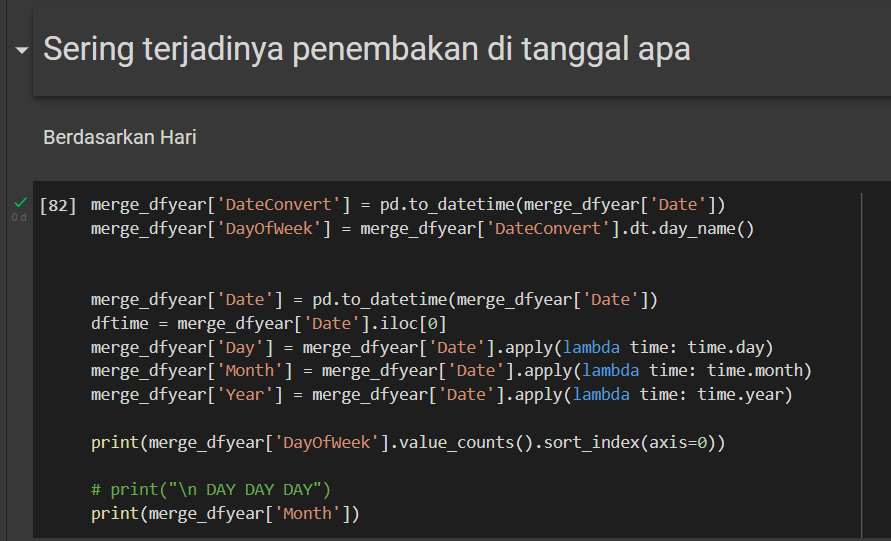
**Gambar 8**



**Gambar 9**

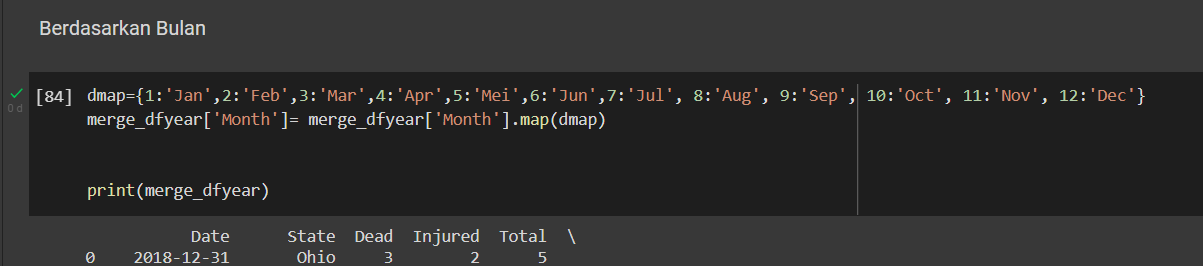
1. Menganalisis kolom ‘Date’

* Mengetahui hari yang paling sering terjadi penembakan



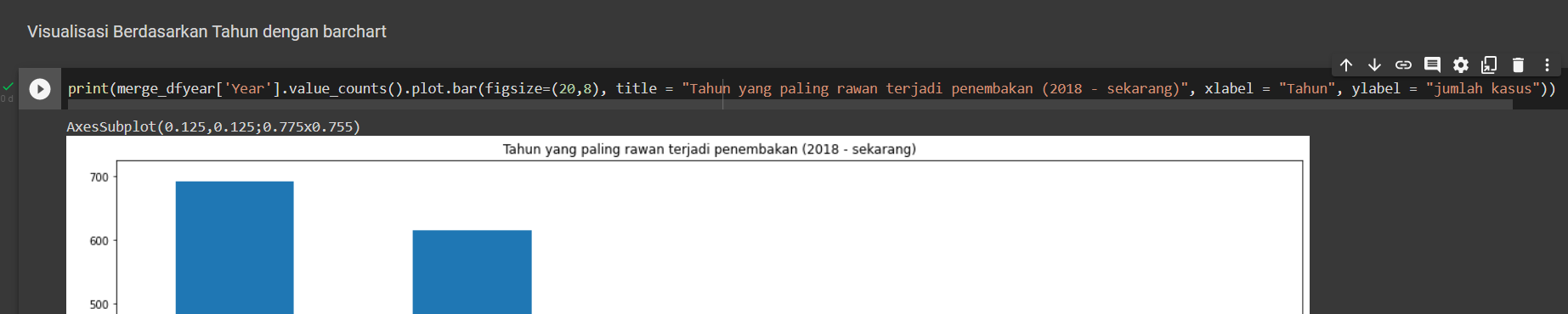
**Gambar 10**

* Mengetahui bulan yang paling sering terjadi penembakan



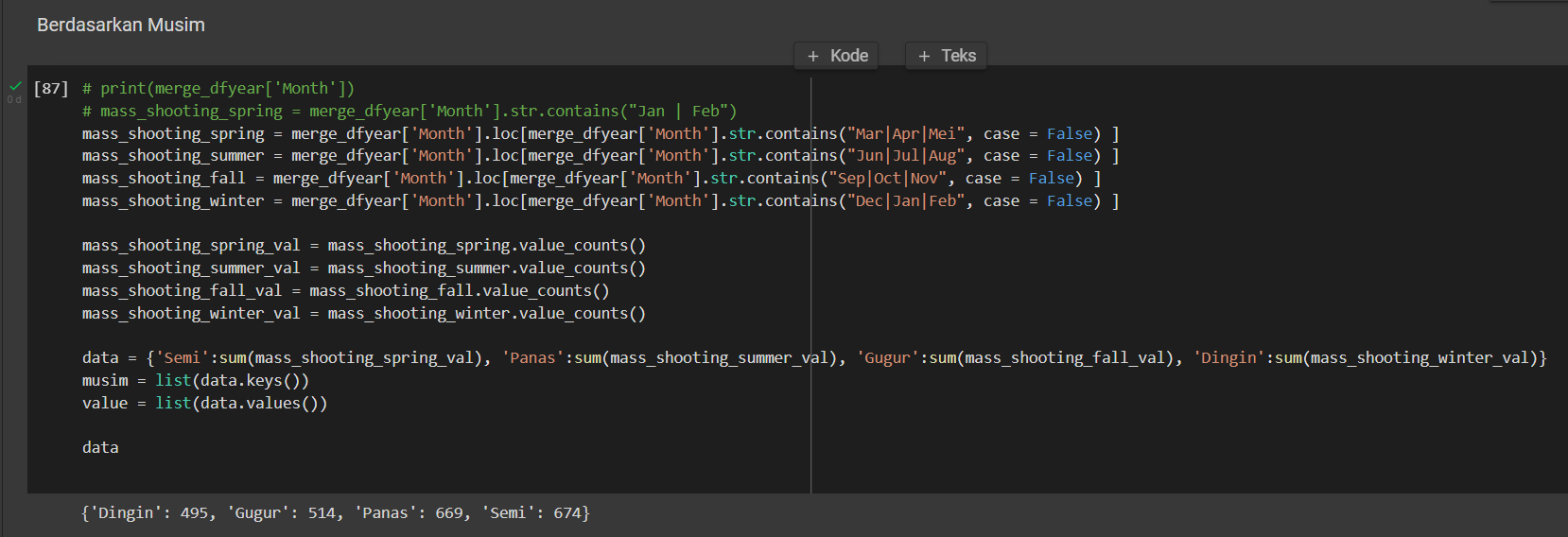
**Gambar 11**

* Mengetahui tahun yang paling sering terjadi penembakan



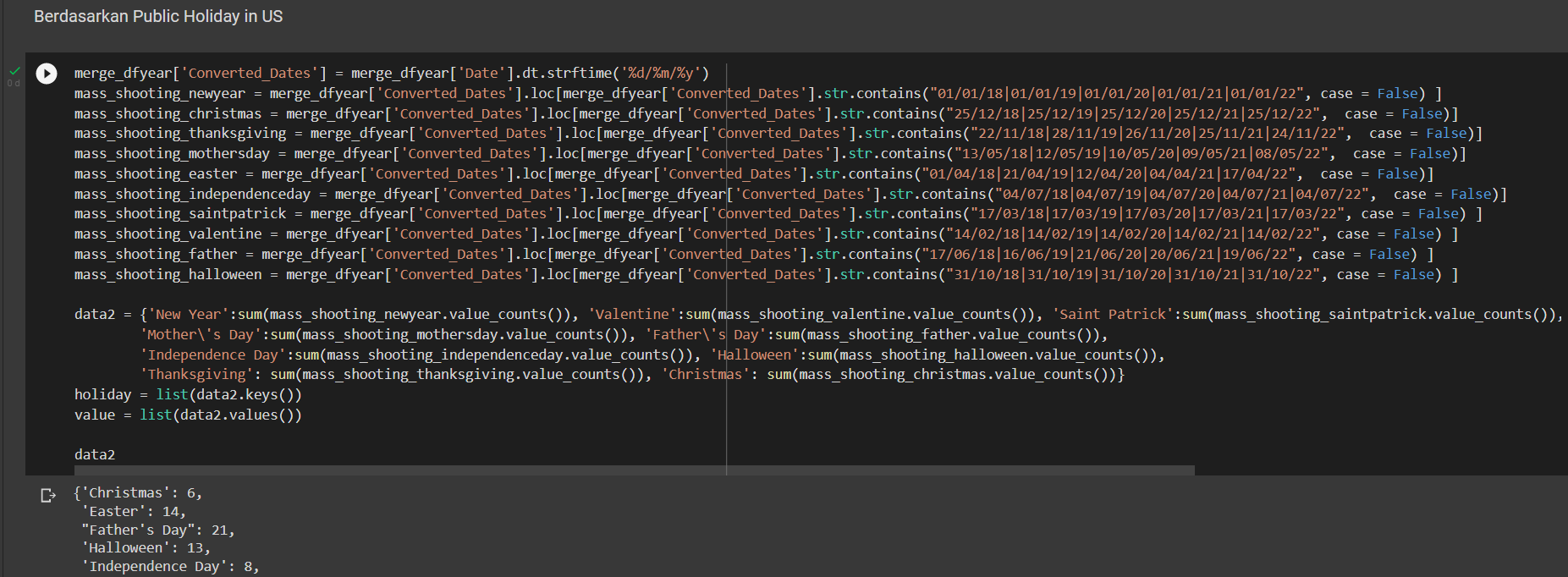
**Gambar 12**

* Mengetahui musim yang paling sering terjadi penembakan



**Gambar 13**

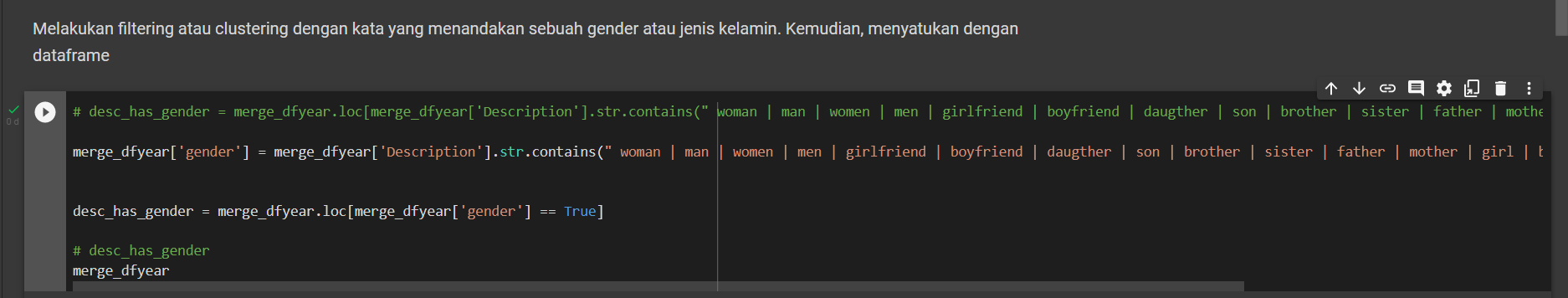
* Mengetahui public holiday yang paling sering terjadi penembakan



**Gambar 14**

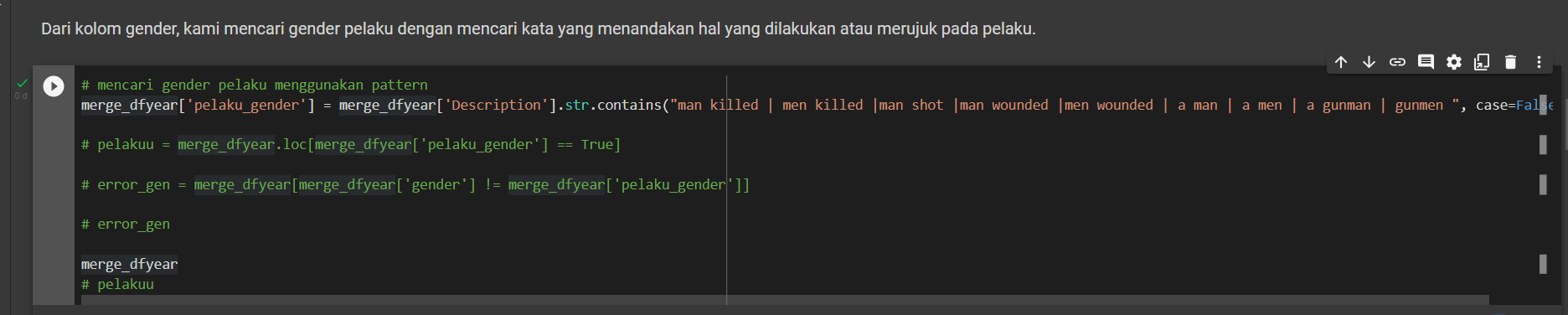
1. Menganalisis kolom ‘Description’

* Mengetahui nilai pada kolom ‘Description’ yang menandakan bahwa gender bisa diketahui dengan cara melakukan filtering kata yang menandakan sebuah gender. Setelah itu, menggabungkan dengan dataframe utama.



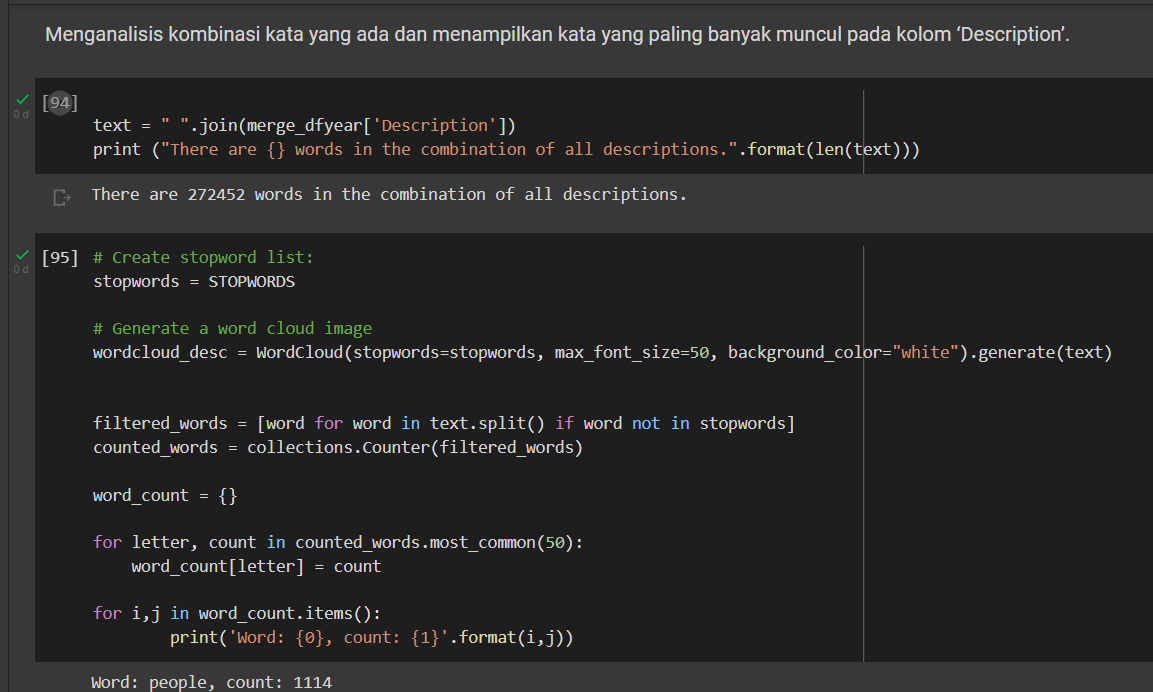
**Gambar 15**

* Mencari gender pelaku dengan melakukan filtering menggunakan kata yang menandakan hal yang dilakukan atau merujuk pada pelaku.



**Gambar 16**

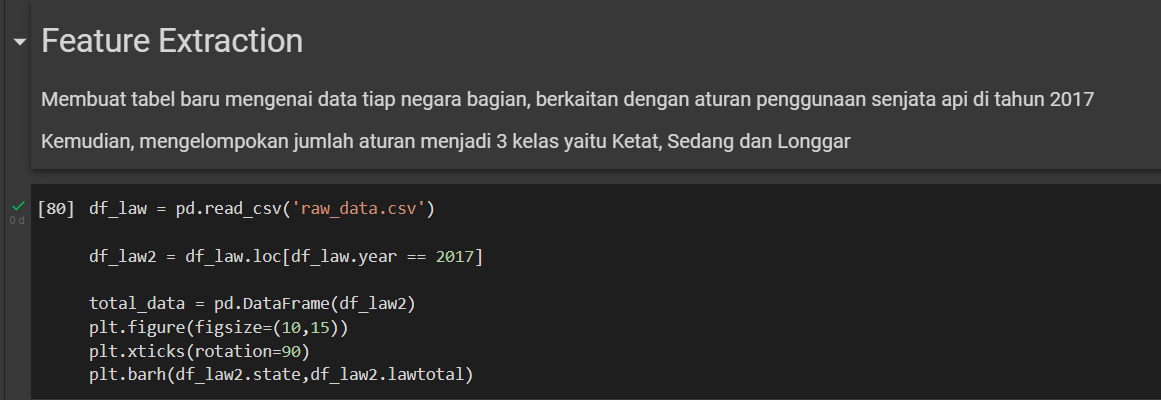
* Menganalisis kombinasi kata yang ada dan menampilkan kata yang paling banyak muncul pada kolom ‘Description’.



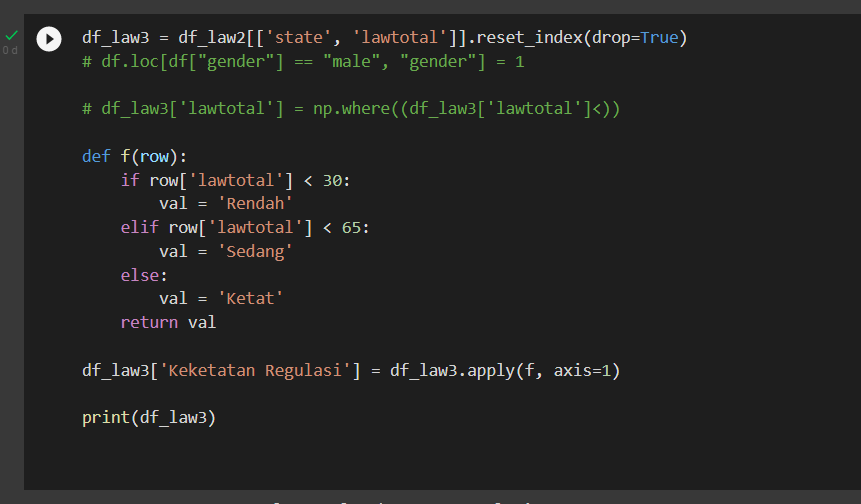
**Gambar 17**

1. Menganalisis aturan atau regulasi penggunaan senjata api di setiap negara bagian

Dengan menggunakan dataset yang baru yaitu mengenai jumlah aturan atau regulasi yang ada di setiap negara bagian Amerika Serikat, kami melakukan analisis mengenai jumlah aturan dan mengelompokan menjadi 3 kelas yaitu Ketat, Sedang dan Longgar



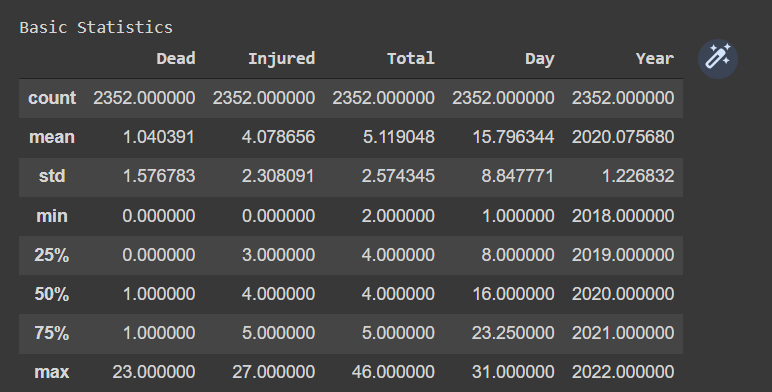
**Gambar 18**



**Gambar 19**

1. **Visualisasi**
   1. Basic statistics

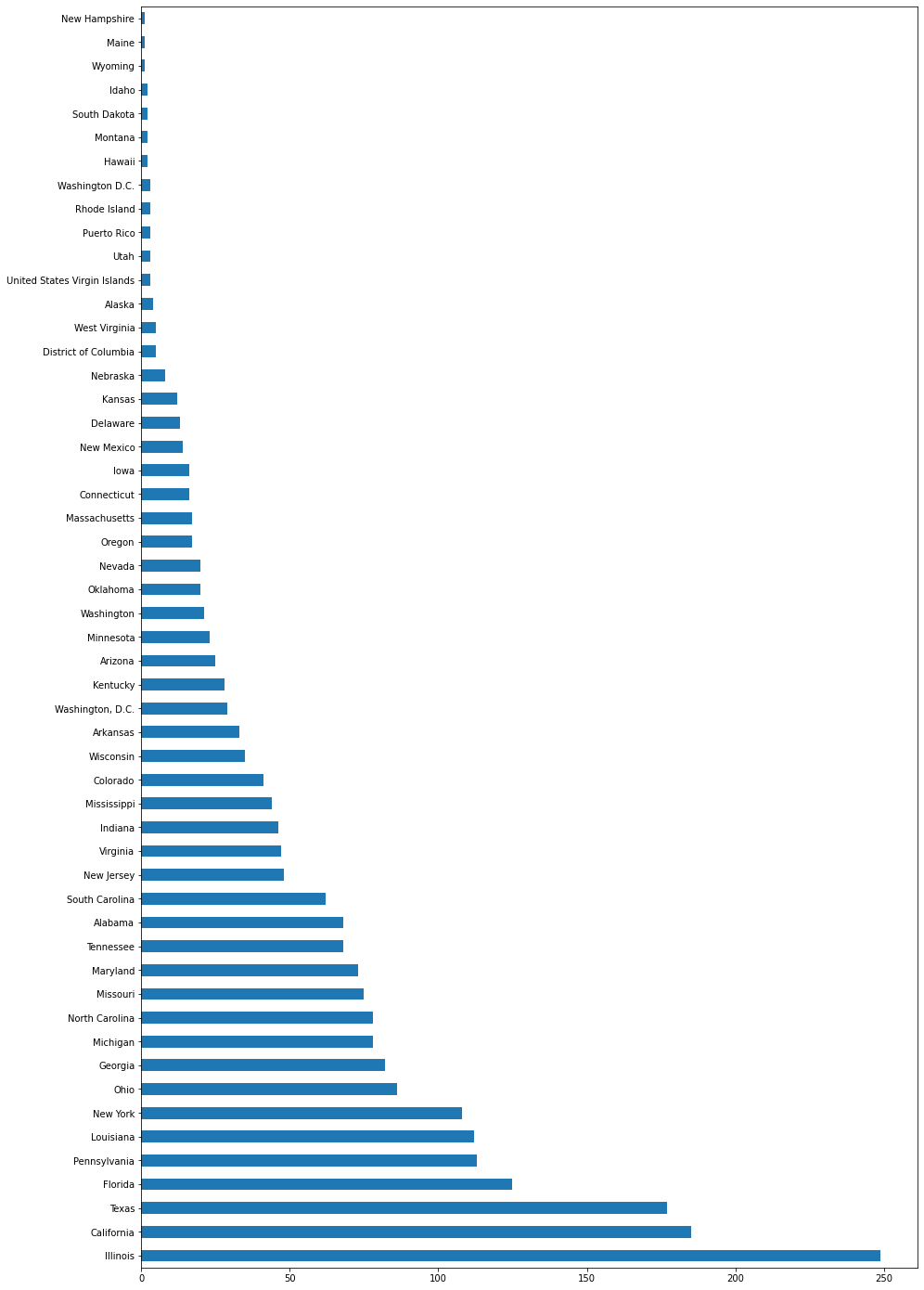
Menggunakan tabel yang dihasilkan oleh functions pandas



**Gambar 20**

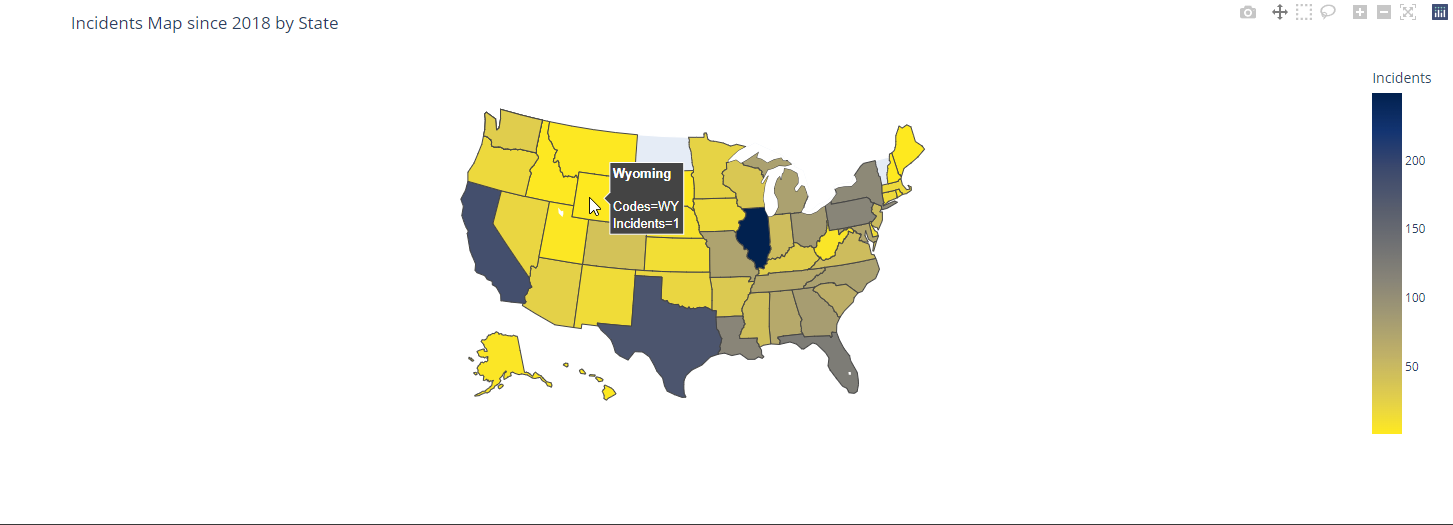
* 1. Total jumlah kasus penembakan yang terjadi di setiap negara bagian

menggunakan bar chart



**Gambar 21**

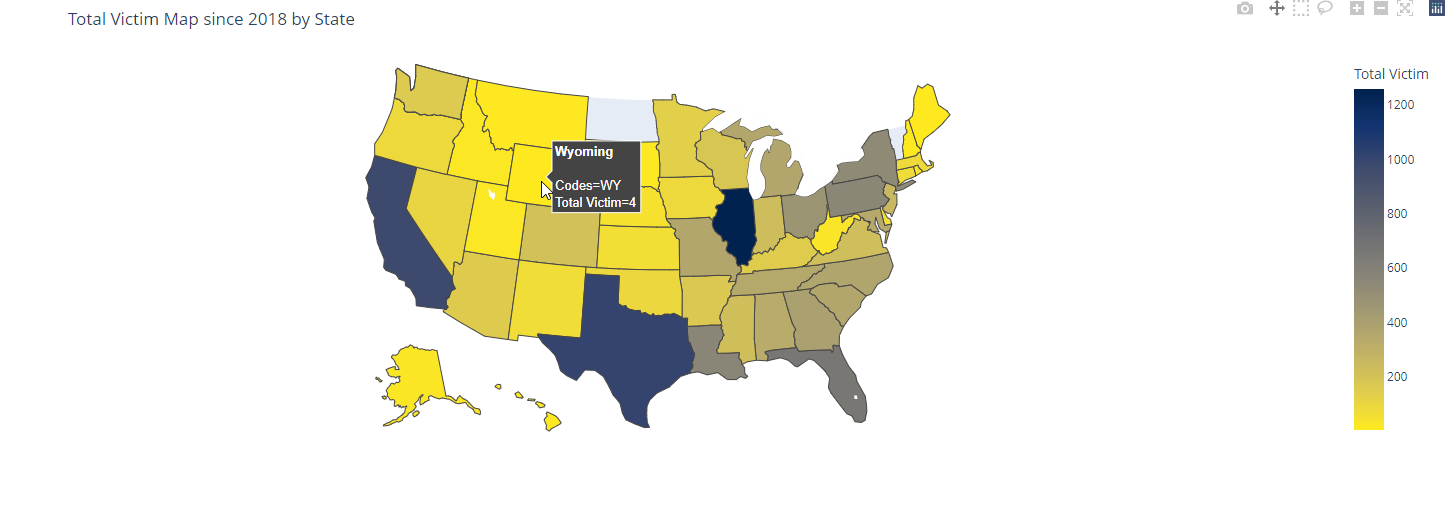
Selain itu, kami memvisualisasikan juga menggunakan Spatial Map



**Gambar 22**

* 1. Total jumlah korban dalam setiap negara bagian

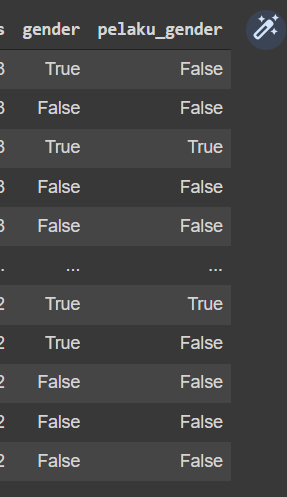
Menggunakan spatial map



**Gambar 23**

* 1. Attribute baru yaitu ‘gender’ dan ‘pelaku\_gender’

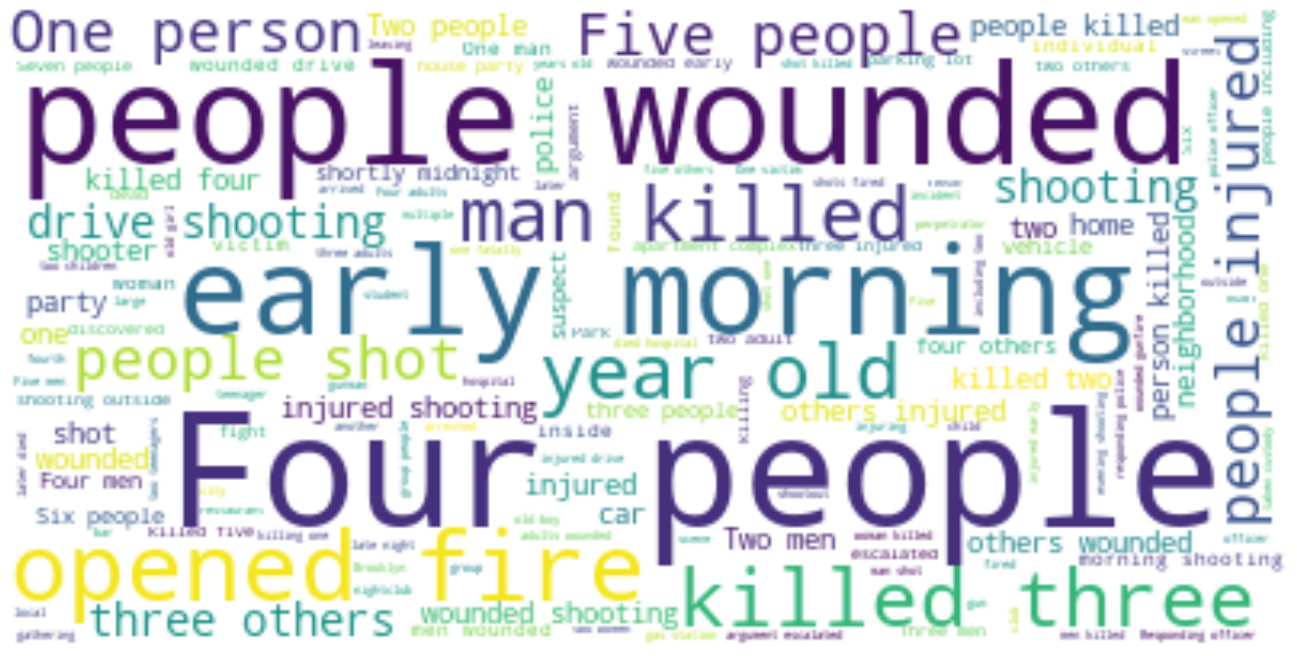
Menambahkan kolom atau attribute baru ke dataframe



**Gambar 24**

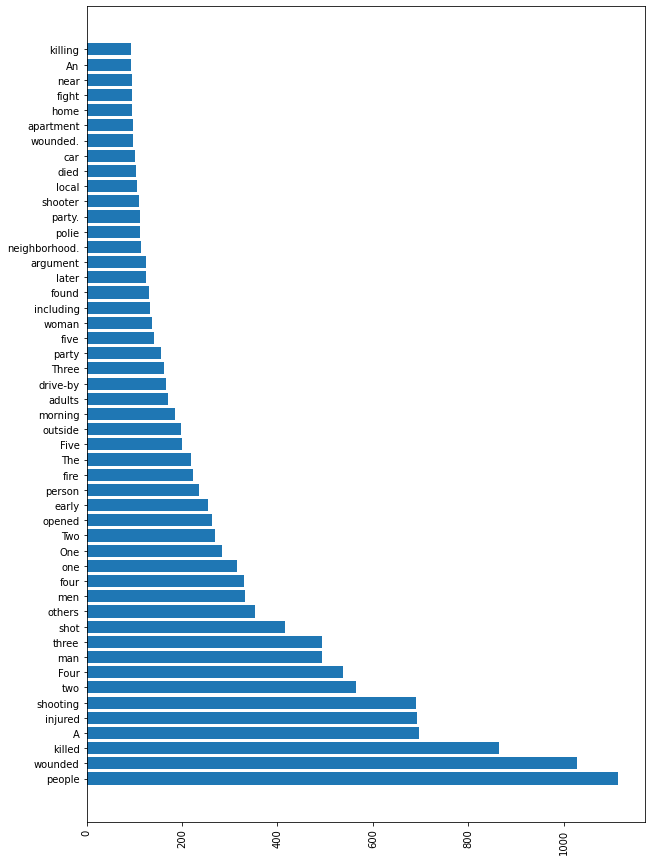
* 1. Kombinasi kata yang paling banyak muncul pada kolom ‘Description’.

Menggunakan wordcloud



**Gambar 25**

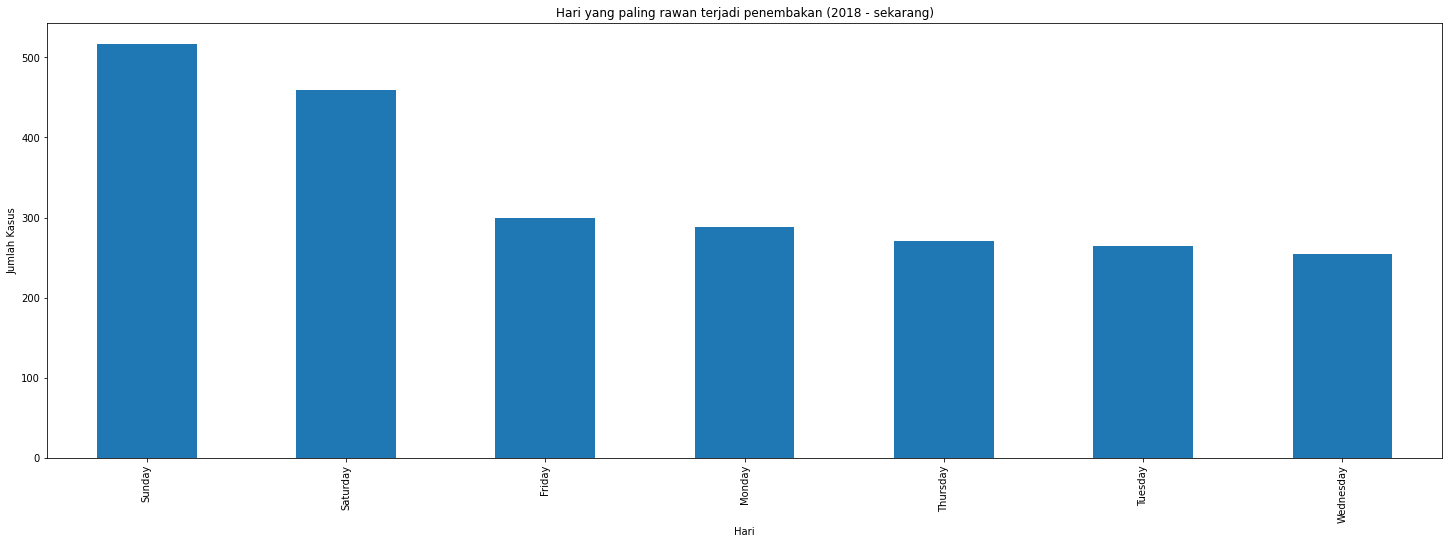
Menggunakan barchart



**Gambar 26**

* 1. Penembakan sering terjadi dikelompokan berdasarkan
* Hari

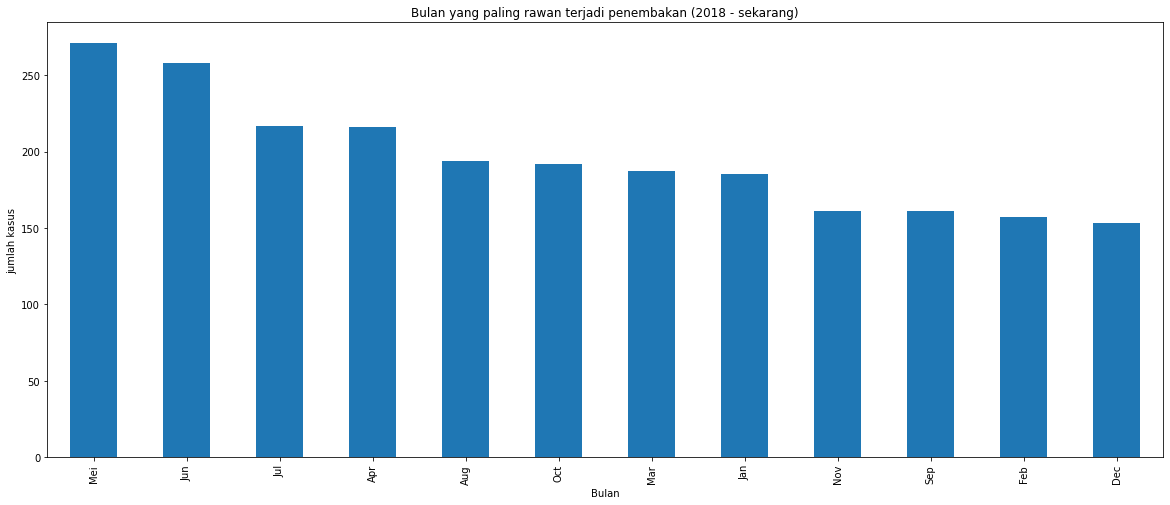
Menggunakan barchart



**Gambar 27**

* Bulan

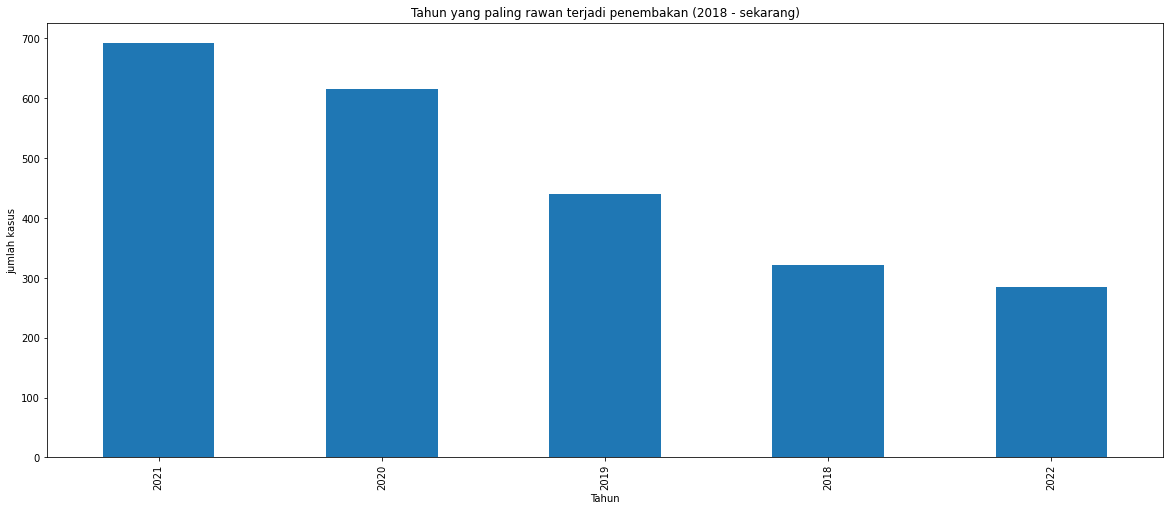
Menggunakan barchart



**Gambar 28**

* Tahun

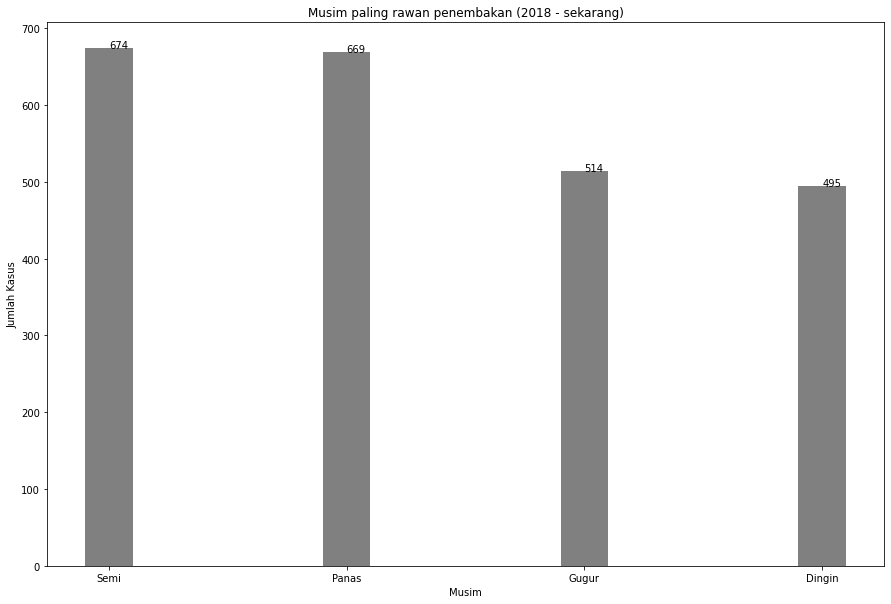
Menggunakan barchart



**Gambar 29**

* Musim

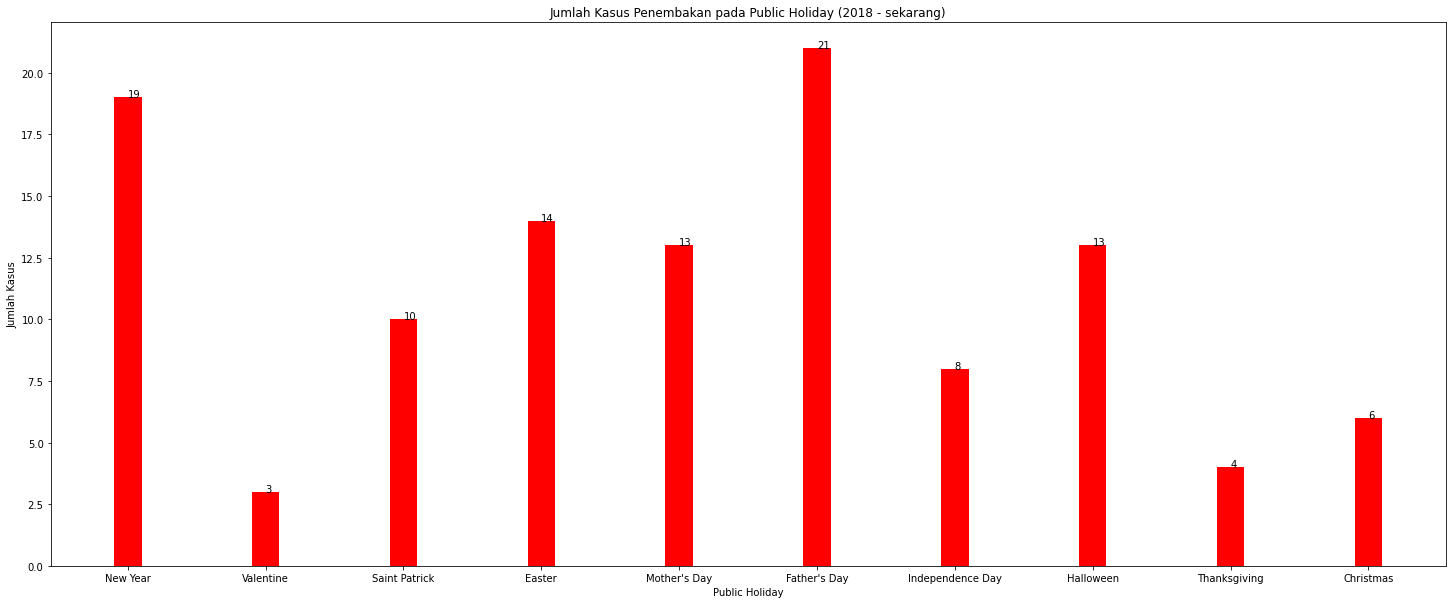
Menggunakan barchart



**Gambar 30**

* Public Holiday

Menggunakan barchart



**Gambar 31**

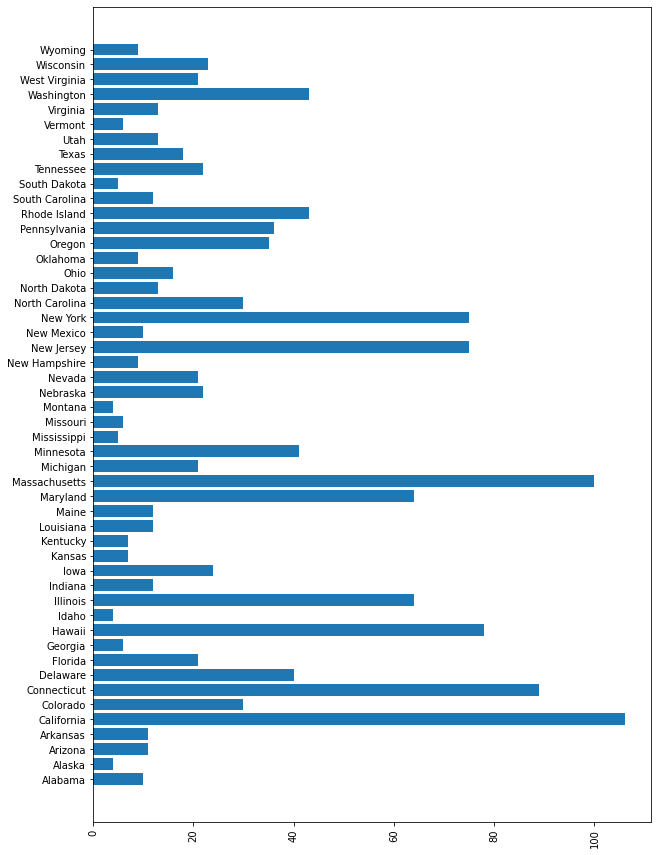
* 1. Aturan atau regulasi penggunaan senjata api di setiap negara bagian dan dikelompokan menjadi 3 kelas yaitu Ketat, Sedang dan Longgar

Menggunakan tabel baru



**Gambar 32**

menggunakan bar chart



**Gambar 33**

1. **Penutup**

Dalam setiap poin analisis yang dilakukan kami mendapatkan, pertama adalah negara bagian Illinois memiliki jumlah kasus penembakan terbanyak dengan jumlah 249 kasus dan jumlah korban terbanyak juga dipegang oleh Illinois sebanyak 1260 korban. Kedua, kami mendapatkan hasil bahwa hari paling banyak terjadi penembakan adalah hari Minggu dan diikuti oleh hari Sabtu. Tetapi dalam data tersebut, didapatkan bahwa hari Jumat memiliki selisih yang cukup banyak dengan hari Sabtu yaitu sebanyak 160 kasus penembakan. Oleh karena itu, kami berpikir bahwa akhir pekan memungkinkan meningkatnya peluang terjadinya penembakan karena banyak penduduk yang tidak menjalani kesibukan apapun dan lebih banyak melakukan aktivitas sosial dengan orang lain.

Selanjutnya, berdasarkan data jumlah penembakan menurut bulan dan tahunnya, bulan Mei memiliki jumlah kasus terbanyak dengan jumlah 271 kasus penembakan dan diikuti oleh bulan Juni. Lalu, untuk kasus jumlah penembakan pertahunnya, tahun dengan jumlah kasus terbanyak adalah tahun 2021 dengan jumlah kasus penembakan sebanyak 692 dan diikuti oleh tahun 2020. Kami berhipotesis bahwa tahun 2020 dan 2021 merupakan tahun dengan jumlah kasus terbanyak karena dipengaruhi oleh kondisi masyarakat saat pandemi COVID-19. Selain bulan dan tahun, musim juga mempengaruhi banyaknya kasus penembakkan dalam negara-negara bagian Amerika Serikat. Analisis kami mendapatkan bahwa kasus penembakan terjadi paling banyak pada musim semi dengan jumlah kasus penembakan sebanyak 674 dan diikuti oleh musim panas. Perbandingan drastis antara musim gugur dan panas juga memberikan hipotesis bahwa libur musim panas berkemungkinan meningkatkan kasus penembakan karena pada musim panas penduduk menjalani aktivitas dengan orang lain.

Ketiga, Selama tahun 2017, kami dapat mengetahui bahwa California memiliki 106 regulasi penggunaan senjata api yang merupakan regulasi yang sangat ketat. Namun cukup ironis jika dilihat dari insiden dan jumlah korban yang terjadi. Berbanding terbalik dengan negara yang memiliki regulasi yang rendah, seperti Alaska dan Idaho yang memiliki regulasi senjata api yang rendah namun jumlah korban dan insiden yang terjadi jauh lebih rendah dibandingkan California. Kesimpulan yang kami dapatkan, tingkat kasus penembakan massal menggunakan senjata api di Amerika Serikat sangat tinggi meskipun sudah ada banyak regulasi mengenai kepemilikan senjata api. Terakhir, dengan menggunakan kolom ‘Description’, kami dapat mengetahui jumlah kombinasi kata yang ada dan dapat diketahuinya gender pelaku dari kolom tersebut. Untuk kedepannya, yang harus dikembangkan dari analisis kami adalah data mengenai regulasi tentang kepemilikan senjata api di setiap negara bagian Amerika Serikat, karena data yang kami gunakan pada saat ini adalah data dari tahun 2017, dan ada kemungkinan bahwa regulasi-regulasi tersebut sudah diperbaharui oleh pemerintah Amerika Serikat.