

Tugas Rekayasa Data

I. Deskripsi Dataset

Dataset ini memiliki informasi terkait serangan teroris dari tahun 1970 hingga 2016. Terdapat informasi semisal mengenai tanggal terjadinya serangan, lokasi negara, anggota kelompok teror mana yang menyerang, hingga persenjataan dan juga korban dalam serangan tersebut.

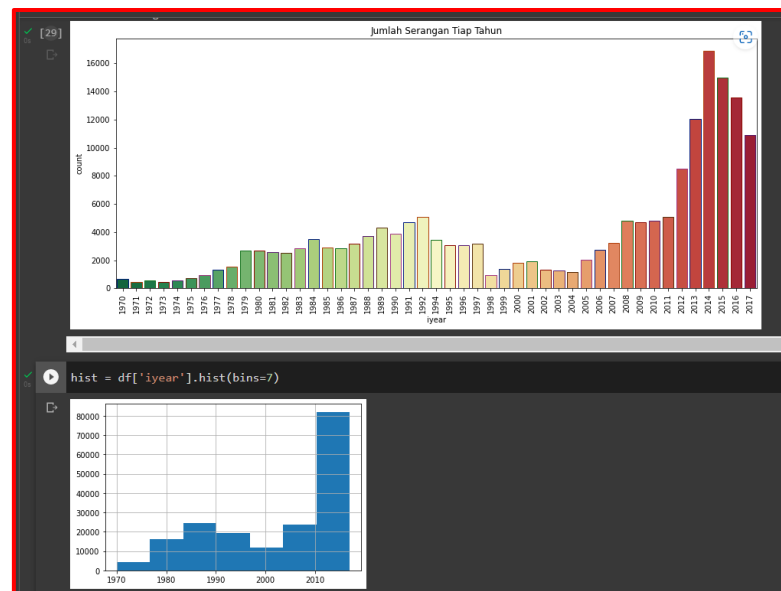
II. Isi

Saya mengambil beberapa bagian dari dataset untuk di visualisasikan dan mengambil analisis, mari ambil contoh jumlah serangan yang terjadi setiap tahunnya.

Berikut informasi numeris yang didapatkan tentang serangan yang terjadi tiap tahunnya.

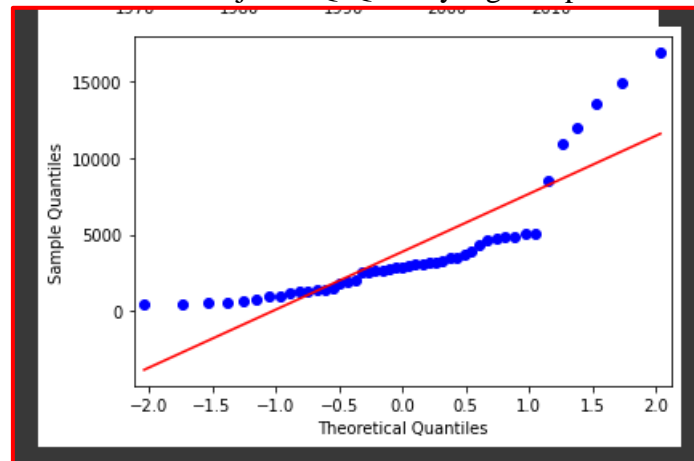
1. Rata-rata serangan = 3865 serangan
2. Median serangan ada di tahun 2009.

Setelah ini kita akan melihat mengapa median di 2009 ini merefleksikan peningkatan serangan yang terjadi.



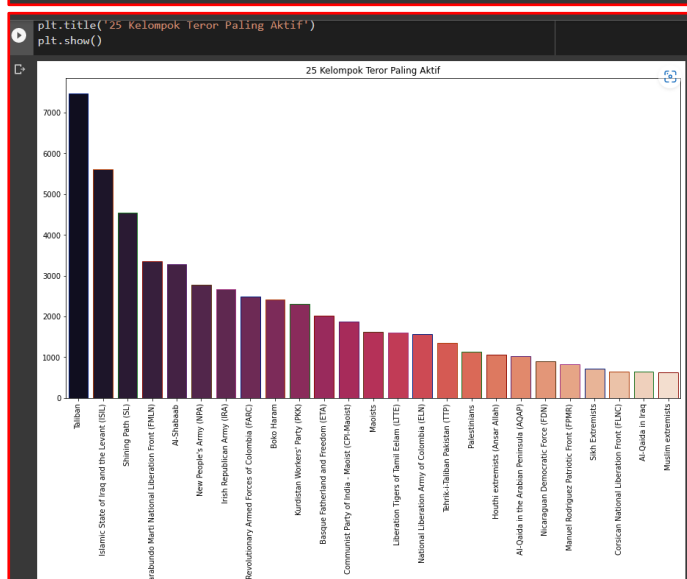
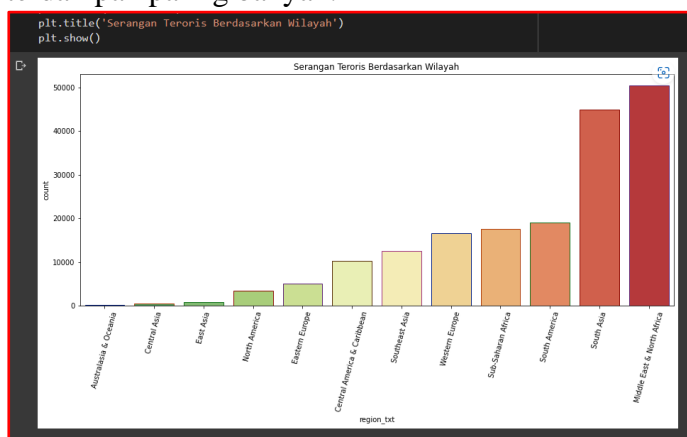
Kita bisa melihat visualisasi dengan menggunakan bar chart biasa (atas) dengan `sns.countplot`, dan juga dengan menggunakan histogram menggunakan library `pandas`. Perbedaannya adalah histogram bisa meringkas informasi karena bagian data akan dibagi kedalam “bins”, jumlah bins bisa kita atur, namun secara umum, semakin banyak data jumlah bins harus disesuaikan untuk ditingkatkan agar data tetap representatif. Histogram tersebut memberitahu kita tentang informasi bahwa diatas tahun 2010, ada sekitar 80,000 serangan yang terjadi, yang memang correspond dengan bar chart yang kita buat yang memang mengalami peningkatan di tahun 2010 keatas.

Selanjutnya kita bisa amati sejenak Q-Q Plot yang didapatkan dari data diatas,



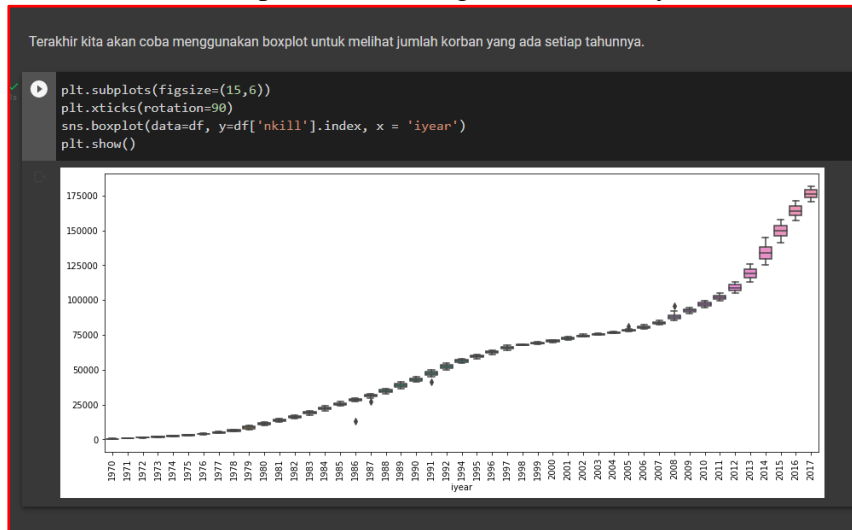
Terlihat bila terbukti jika memang serangan teroris berdasarkan kuantil tahun tidak terdistribusi normal^[1], dan justru cenderung *skewed* dengan drastis semakin ke kanan.

Selanjutnya mari amati korelasi antara kelompok teror yang aktif dengan daerah yang terdampak paling banyak.



Terlihat bahwa serangan yang paling banyak terjadi di Timur Tengah & Afrika, setelah itu jika dibandingkan dengan kelompok paling aktif, yakni Taliban dan ISIL, hal ini memang membuktikan adanya korelasi antara daerah serangan dan juga kelompok yang

beroperasi. Namun hal ini tentu bukan 100% penyebab mengapa di Timur tengah dan afrika terdapat sangat banyak serangan teror.



Dari boxplot ini kita melihat jika jumlah korban selalu meningkat dengan paling banyak berada di 2017. Boxplot menggambarkan jika distribusi korban berdasarkan kuartil relatif kecil di tahun 1970-an hingga awal 2000. Barulah ketika 2012-2017 distribusinya meningkat.

III. Kesimpulan

Kesimpulan saya mengenai 2 bagian dataset yang saya amati adalah yakni,

1. Terkait peningkatan drastis serangan teror di tahun 2000-an, besar kemungkinan penyebabnya adalah
 - a. Motivasi anggota kelompok Jihad dan teror meningkat setelah melihat kesuksesan serangan 11 September di Amerika.
 - b. Invasi ke beberapa negara Timur tengah seperti Iraq dan Afghanistan oleh Koalisi amerika dan eropa
2. Terkait mengapa negara Timur tengah paling sering mengalami serangan teroris, dan merupakan persebaran utama kejadiannya, penyebabnya kemungkinan adalah,
 - a. Maraknya perang sipil yang terjadi di daerah seperti Suriah & Yaman.
 - b. Tidak adanya penegakan hukum yang menyebabkan kelompok seperti Al Qaeda dan ISIS untuk tetap beroperasi.

Referensi

[1] <https://towardsdatascience.com/q-q-plots-explained-5aa8495426c0#:~:text=Q%2DQ%20plots%20are%20used%20to,or%20even%20Pareto%20Distribution%2C%20etc.>

Link colab :

<https://colab.research.google.com/drive/1JOlhyXTrGlaJXPSNTAOFrRgjP7bakQy0?usp=sharing>