

Deskripsi Data

(Link sumber ada di grup, dapatkan penjelasan dari link sumber tersebut)

Data ini menunjukkan produksi listrik oleh negara-negara di dunia yang dikategorisasikan dalam beberapa sumber seperti minyak, surya, dan nuklir (lengkapnya di bagian selanjutnya) dari tahun ke tahun (spesifikasikan di laporan kalau tahunnya dari 1985 sampai 2020, akan tetapi beberapa negara hanya memiliki data tahun 2000 ke atas dan beberapa negara hanya punya sampai 2019). Dari data ini bisa didapatkan tren sumber tenaga yang dipakai semua negara di dunia, ataupun salah satu negara, seiring berjalannya waktu.

12 kolom dan 6242 baris

Karakteristik Data

Kolom produksi listrik adalah atribut kuantitatif dengan range ≥ 0 . Nama negara, dan kode mereka adalah atribut kategorikal-nominal.

Statistik Data

Data yang diambil statistiknya adalah produksi listrik tiap negara pada tahun 2019. Kenapa kami mengambil satu tahun tertentu dan bukan semua tahun saja? Karena menurut kami mengiris satu periode waktu untuk diperiksa lebih masuk akal daripada menganalisis seluruh data secara keseluruhan. *Insight* yang didapatkan akan lebih terfokus dan mendalam.

Visualisasi Data

Hitam digunakan untuk bahan bakar fosil, hijau digunakan untuk bahan bakar terbarukan, dan kuning digunakan untuk energi nuklir karena mereka telah terasosiasi dengan warna-warna tersebut (Secara intuisi sudah jelas tapi aku nggak tahu cara menjelaskannya secara eksak, jelaskan lebih lengkap). (Untuk warna yang aku pilih untuk masing-masing tipe seperti surya atau nuklir udah jelas kan ya. Bisa dikarang sendiri penjelasannya).

- Perbandingan kategori :
 - Perbandingan sumber-sumber produksi di USA tahun 2019
 - Perbandingan sumber-sumber produksi di Indonesia tahun 2019
- Penampilan perubahan terhadap waktu :
 - Pertumbuhan sumber energi fosil, terbarukan, dan nuklir yang dipakai Irlandia dari 2000-2019
 - Pertumbuhan sumber energi fosil, terbarukan, dan nuklir yang dipakai Jepang dari 2000-2019
- Penampilan hierarki dan hubungan keseluruhan-bagian :

- Menampilkan dari total energi yang dihasilkan Prancis tahun 2000-2019, berapa yang berasal dari nuklir, berapa yang berasal dari minyak, dan seterusnya.
- Menampilkan dari total energi yang dihasilkan Jerman tahun 2000-2019, berapa yang berasal dari nuklir, berapa yang berasal dari minyak, dan seterusnya.
- *Plotting relationships* :
 - Scatter plot dari energi terbarukan vs energi fosil seiring berjalannya waktu dari negara Belgia
 - Scatter plot dari energi terbarukan vs energi fosil seiring berjalannya waktu dari negara India

Korelasi Data

Kami menggabungkan atribut-atribut yang ada menjadi 3, yaitu bahan bakar fosil (batu bara, minyak, dan gas), energi terbarukan (surya, hidro, angin, dan terbarukan lainnya), dan nuklir. Kami kemudian melihat korelasi antar ketiganya.

Data Cleansing

Walaupun data yang ada di sini mencakup tahun 1985-2020, tidak semua negara memiliki data untuk tahun 1985-1999 dalam artian tidak ada baris untuk tahun tersebut. Terlebih lagi, negara-negara yang memiliki baris tahun-tahun tersebut, banyak yang beberapa kolom produksinya kosong di tahun itu. Hal yang sama juga terjadi pada tahun 2020, tidak semua negara memiliki baris tahun 2020. Kami memutuskan untuk membuang baris 1985-1999 dan 2020 yang ada demi penyeragaman dan menghindari kesalahan karena mencoba menjumlahkan tipe data NaN.

Dari total 235 negara dan entitas supranasional seperti benua ataupun kumpulan negara yang ada di data, hanya 96 yang memiliki baris tahun 1985-1999. Dari 96 negara tersebut, hanya yang semua kolom produksinya lengkap dan tidak ada yang kosong pada baris tahun tersebut. Hanya 71 yang memiliki baris tahun 2020. Dari 71 tersebut hanya yang semua kolom produksinya lengkap dan tidak ada yang kosong pada baris 2020. (Kalkulasikan baris yang tidak bisa dipakai lalu sebutkan berapa).

Kekosongan data tidak hanya terjadi pada tahun 1985-1999, tetapi juga pada tahun 2000 ke atas. Setelah proses penghapusan baris 1985-1999 dan 2020, kami memutuskan untuk membuang negara dan entitas yang tidak lengkap datanya pada tahun 2000-2019 karena tidak mungkin mengolah data yang tidak lengkap, apalagi membuat sebuah kesimpulan yang akurat dari sana.

Untuk pembersihan data, kami memutuskan untuk tidak memodifikasi file orisinalnya, tetapi membuat file baru yang sudah dibersihkan, ini dilakukan agar jika terdapat kesalahan dalam proses pembersihan, kedua data bisa dibandingkan untuk mengetahui di mana kesalahannya dan prosesnya segera dibetulkan.

