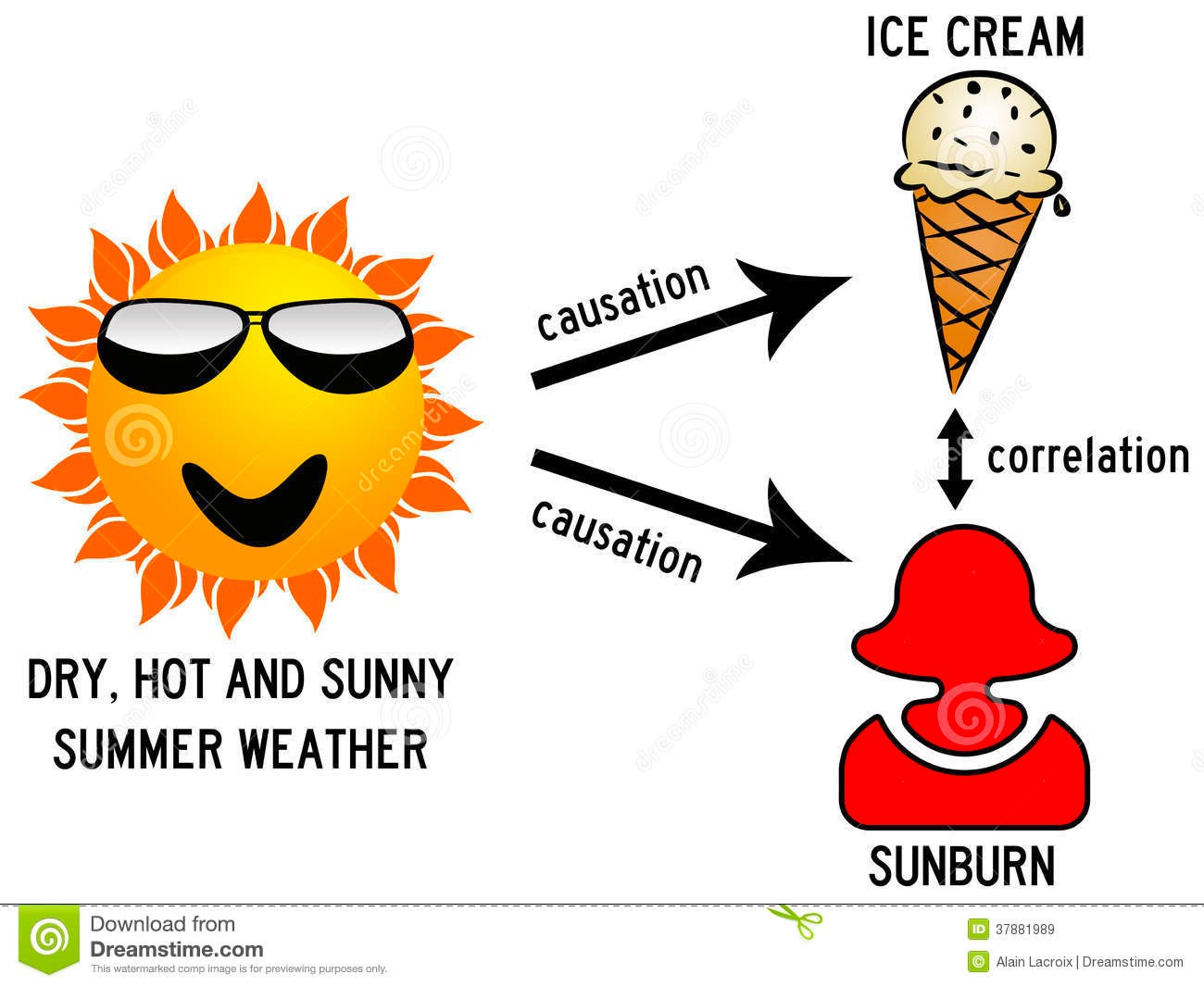
1. Jelaskan secara teori statistik mengenai outlier (pencilan), implikasinya dalam analisis data, serta bagaimana melakukan manajemen data terhadap kasus outlier.
   * Menurut *Engineering Statistics Handbook[[1]](#footnote-0),* Outlier adalah data yang letaknya berada pada jarak yang abnormal dari data-data lain. Sedangkan menurut Moore and McCabe dalam John[[2]](#footnote-1), Outlier adalah data yang terletak diluar pola keseluruhan dari sebuah distribusi. Sehingga outlier adalah data yang nilainya memiliki perbedaan yang signifikan dengan mayoritas data yang ada. Cara menentukan apakah sebuah data adalah outlier ada dua, yaitu berdasarkan standar deviasi dan kuartil.
   * Apabila kita ingin memakai data yang ada untuk membuat model machine learning, ada baiknya data outlier dibuang sehingga model yang dibuat nanti memiliki performa yang baik dan tidak terpengaruh dengan adanya *outlier*.
2. Jelaskan konsep dan prinsip korelasi, lalu sebisa mungkin kaitkan dengan dasar-dasar statistik serta implikasinya terhadap konsep/teori statistik lain.
   * Korelasi adalah sebuah teknik dalam statistika untuk menentukan kuat pengaruh satu variabel dengan variabel lain. Korelasi antara dua variabel semakin berbanding lurus secara linear (semakin besar variabel A, semakin besar variabel B) bila nilai korelasi mendekati satu. Bila nilai korelasi semakin mendekati negatif satu, maka hubungan antara dua variabel akan semakin berbanding terbalik secara linear (semakin besar variabel A, semakin kecil variabel B).



Gambar 1. Penggambaran korelasi dan kausalitas.

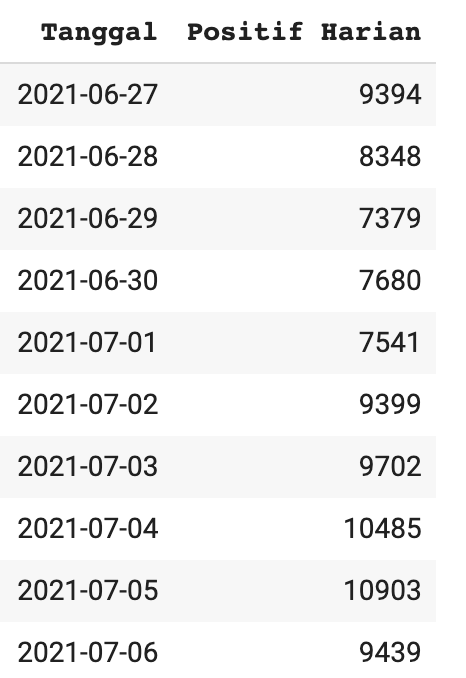
Korelasi memang menggambarkan hubungan antara satu variabel dengan variabel lain, namun tidak selalu menggambarkan hubungan yang terjadi karena keterkaitan suatu hal atau kausalitas. Semisal bila dilihat dari korelasi, kenaikan kasus kulit terbakar (*sunburn*) berbanding lurus dengan penjualan es krim. Semakin banyak es krim yang terjual, maka semakin tinggi juga kasus kulit terbakar. Namun dibalik korelasi tersebut ada variabel lain yang juga berperan penting yaitu suhu dan tingkat konsentrasi sinar UV. Semakin tinggi suhu harian, semakin tinggi juga penjualan es krim. Hal itu dapat kita logikakan dengan sederhana. Begitu juga dengan konsentrasi sinar UV harian, semakin tinggi suhu pada hari tersebut, konsentrasi sinar UV akan semakin tinggi dan dapat menyebabkan *sunburn* [[3]](#footnote-2).

1. Sebutkan teori dasar machine learning yang kalian ketahui, lalu jelaskan dalam bahasa sederhana mengenai teori tersebut dan implikasinya.
   * Menurut Tom Mitchell, pembelajaran mesin adalah

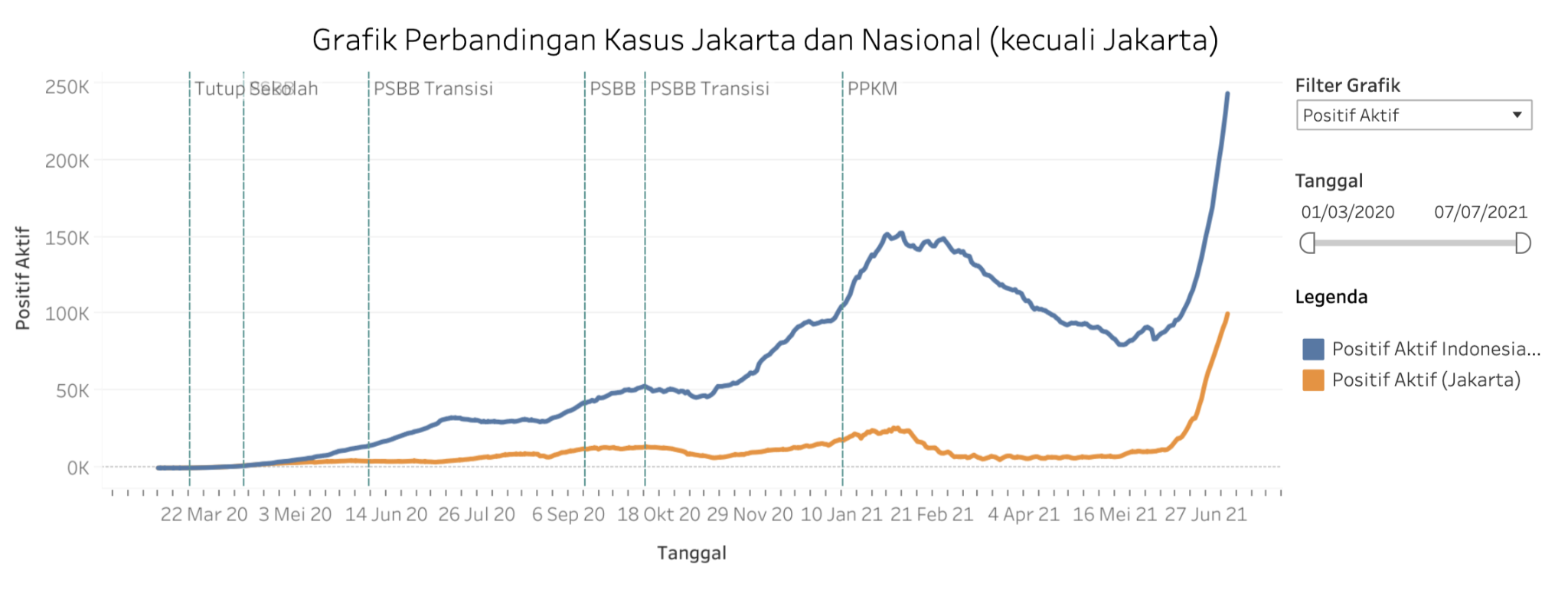
*“Sebuah komputer di program untuk belajar dari pengalaman (data) E yang berkaitan dengan beberapa tugas (class) T dan ukuran performansi P saat melaksanakan tugas T, saat diukur performa P meningkat dengan pengalaman E” [[4]](#footnote-3).*

* + Machine learning adalah suatu sistem yang dibuat untuk menyelesaikan permasalahan dengan cara mempelajari data-data yang berkaitan agar mendapatkan pengalaman guna menyelesaikan permasalahan tersebut.
  + Pada machine learning, terdapat konsep *garbage in garbage out,* yaitu sebaik apapun model algoritma yang digunakan, apabila data yang digunakan itu kualitasnya buruk maka model yang dihasilkan akan linear dengan kualitas data yang diberikan. Sehingga penting untuk seorang *machine learning engineer* untuk memahami data yang diperlukan guna membuat model yang baik.

1. Menggunakan bahasa kalian sendiri, jelaskan kaitan antara artificial intelligence, machine learning, dan deep learning.
   * Artificial Intelligence adalah kemampuan program untuk belajar dan berpikir seperti manusia sehingga dapat menyelesaikan suatu permasalahan tertentu.
   * Machine Learning adalah sebuah algoritma yang memiliki kemampuan untuk belajar tanpa perlu di program secara eksplisit dengan cara dilatih menggunakan data yang bersangkutan dengan permasalahan yang ingin diselesaikan.
   * Deep Learning adalah bagian dari machine learning yang dimana algoritma ini berbentuk seperti jaringan syaraf tiruan yang dapat beradaptasi dan belajar dari data dalam jumlah yang besar.
2. Apakah yang kalian ketahui mengenai interpretasi data? Bagaimana signifikansi dan tantangannya? Bagaimana kaitan interpretasi data dengan *data story telling* dan *decision making*?
   * Interpretasi data adalah suatu metode yang digunakan untuk menyampaikan arti atau temuan yang terdapat pada data, pada umumnya seorang *data analyst* akan melakukan interpretasi data dengan cara visualisasi sehingga data dapat dengan lebih mudah untuk dipahami oleh *stakeholder.* Proses visualisasi ini sangatlah penting dikarenakan kita sebagai manusia akan lebih mudah untuk memahami sesuatu berdasarkan visual.



Gambar 2. data tanpa visualisasi, susah untuk melihat pola di dalamnya.



Gambar 3. Dengan visualisasi data, kita dapat memahami pola dengan lebih mudah.

Kaitan interpretasi data dengan *data story telling* adalah dalam melakukan *data storytelling,* dibutuhkan informasi yang bermanfaat, memiliki tujuan dan cerita yang jelas, dan juga terdapat visualisasi yang menarik sehingga orang yang melihat akan dengan cepat memahami data yang ada.

Sedangkan kaitan dari interpretasi data dengan decision making adalah dengan adanya interpretasi data yang informatif dan mencantumkan juga rekomendasi aksi selanjutnya yang dapat diambil, orang yang melihat interpretasi data tersebut dapat dengan yakin mengambil keputusan dikarenakan telah melihat situasi yang ada dari data dan rekomendasi keputusan yang perlu diambil.

1. NIST, U.S. Department of Commerce. 2012, *Engineering Statistics Handbook.* [*https://www.itl.nist.gov/div898/handbook/prc/section1/prc16.htm*](https://www.itl.nist.gov/div898/handbook/prc/section1/prc16.htm) [↑](#footnote-ref-0)
2. Renze, John. "Outlier." dari MathWorld--A Wolfram Web Resource, oleh Eric W. Weisstein. [*https://mathworld.wolfram.com/Outlier.html*](https://mathworld.wolfram.com/Outlier.html) [↑](#footnote-ref-1)
3. Sulistiya, Prima. 2020. *Beda Korelasi dan Kausalitas, Agar Tak Salah Ambil Kesimpulan.* Tersedia pada <https://mojok.co/prm/komen/versus/beda-korelasi-dan-kausalitas-agar-tak-salah-ambil-kesimpulan/> [↑](#footnote-ref-2)
4. Sahil, Verma. 2018. *Machine Learning Notes 1.* Tersedia pada <https://hackernoon.com/machine-learning-notes-1-c496591f49b8> [↑](#footnote-ref-3)