**Pengenalan Analisis Data**

Salah satu keahlian yang perlu dimiliki oleh seorang data scientist tentunya adalah menganalisis sebuah data hingga menjadi informasi yang *insightful.*Maka dari itu, pada modul ini Anda akan menjelajahi ilmu analisis data hingga menjalankan latihan di dalamnya.

**Berkenalan dengan Analisis Data**

Analisis data adalah proses menyelidiki, membersihkan, mengubah, dan memodelkan data dengan tujuan menemukan informasi yang berguna, menginformasikan kesimpulan, mendukung pengambilan keputusan, dan masih banyak lagi.

Dalam sebuah penelitian tentunya seseorang akan membutuhkan analisis data. Dengan adanya kegiatan tersebut, seseorang dapat lebih mudah untuk mengolah data menjadi sumber informasi baru yang akurat dan terpercaya. Namun, pernahkah terlintas di pikiran Anda, “Mengapa analisis data begitu penting?” atau pernahkah terlintas “Bagaimana ya kalau tidak melakukan analisis data?”.

**Mengapa Data Begitu Penting?**

Dalam era teknologi ini, masuknya gelombang data telah menjadikan setiap perusahaan berlomba-lomba dalam memanfaatkan informasi tersebut menjadi wawasan yang berharga untuk kelangsungan bisnis mereka. Data dianggap sebagai aset berharga yang dapat memberikan keuntungan kompetitif. Analisis data menjadi langkah penting karena membantu para pebisnis mengoptimalkan kinerja mereka.

Menerapkan analisis data ke dalam model bisnis menjadi sangat signifikan karena memungkinkan kita untuk merancang strategi berdasarkan bukti yang tercatat, menargetkan pemasaran berdasarkan perilaku pelanggan, dan meningkatkan produktivitas secara keseluruhan. Dengan data yang akurat dan dikelola dengan baik, perusahaan dapat membuat keputusan yang lebih tepat dan responsif terhadap perubahan pasar.

Selain itu, ungkapan **"*Data is The New Oil*"**semakin mencerminkan nilai pentingnya data. Sebagaimana minyak memiliki peran sentral dalam mendukung industri dan ekonomi, data menjadi landasan bagi inovasi, pengambilan keputusan yang cerdas, dan pengembangan strategi yang efektif. Kesadaran akan nilai data ini semakin memotivasi perusahaan untuk mengumpulkan, mengelola, dan menganalisis data dengan seksama untuk meraih kesuksesan dalam lingkungan bisnis yang kompetitif.

Berikut merupakan peran analisis data di beberapa sektor.

1. **Media dan Hiburan**

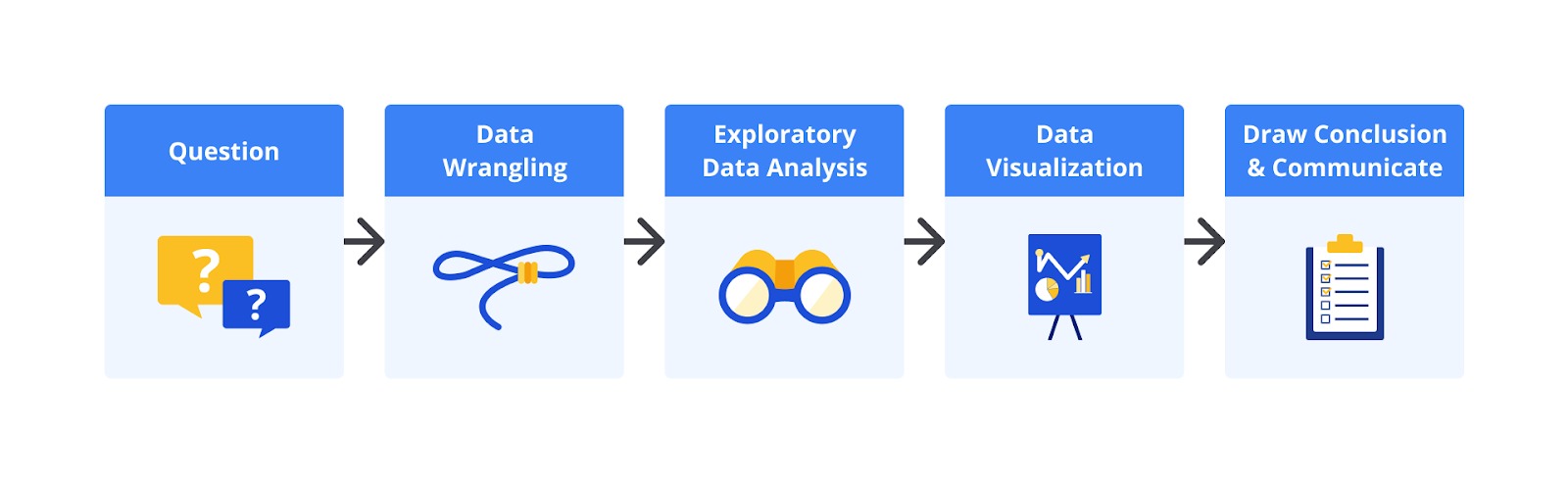
Siapa *sih*yang tak kenal dengan Spotify? Salah satu perusahaan media dan hiburan yang sangat banyak penggunanya di dunia. Nah, perusahaan penyedia musik tersebut merupakan salah satu yang menerapkan analisis data untuk mengumpulkan dan menganalisis data dari jutaan penggunanya sehingga dapat memberikan rekomendasi musik yang lebih baik.

1. **Retail dan Wholesale**

Industri ini memiliki persaingan yang sangat tinggi dan memiliki tantangan tersendiri dalam hal pengelolaan datanya. Para pebisnis menggunakan analisis data untuk menangani data yang besar untuk menangani kebutuhan pelanggan secara akurat. Contoh perusahaan yang menggunakan analisis data di sektor retail dan wholesale adalah Walmart.  
  
Sebagian dari Anda mungkin baru mendengar nama Walmart. Perusahaan waralaba tersebut merupakan perusahaan retail terbesar dan memiliki pendapatan yang tidak main-main. Sejak 2012, Walmart telah memanfaatkan analisis *big data*untuk kelangsungan perusahaannya sehingga perusahaan tersebut dapat mengetahui secara detail tentang konsumen mereka.

**Siklus Analisis Data**

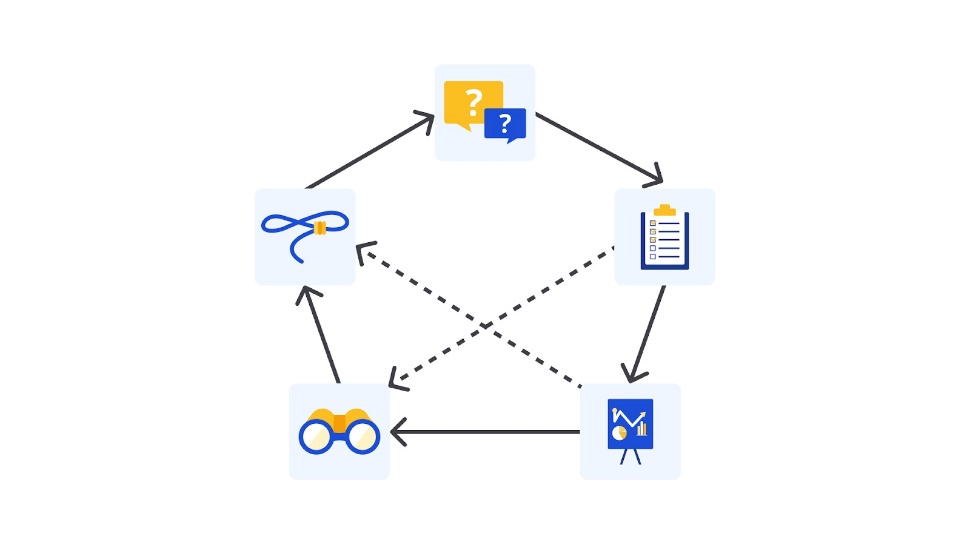
Siklus analisis data umumnya seperti yang tertera di bawah ini. Namun, perlu *disclaimer*di awal bahwa pola tersebut bukanlah *waterfall*alias setiap tahap harus berurutan secara mutlak, tetapi bisa jadi Anda menemukan kondisi di mana tahapan dalam analisis data tidak seperti di bawah ini dan Anda bisa menyesuaikannya dengan kebutuhan. Simak penjelasan siklus analisis data berikut dengan saksama.



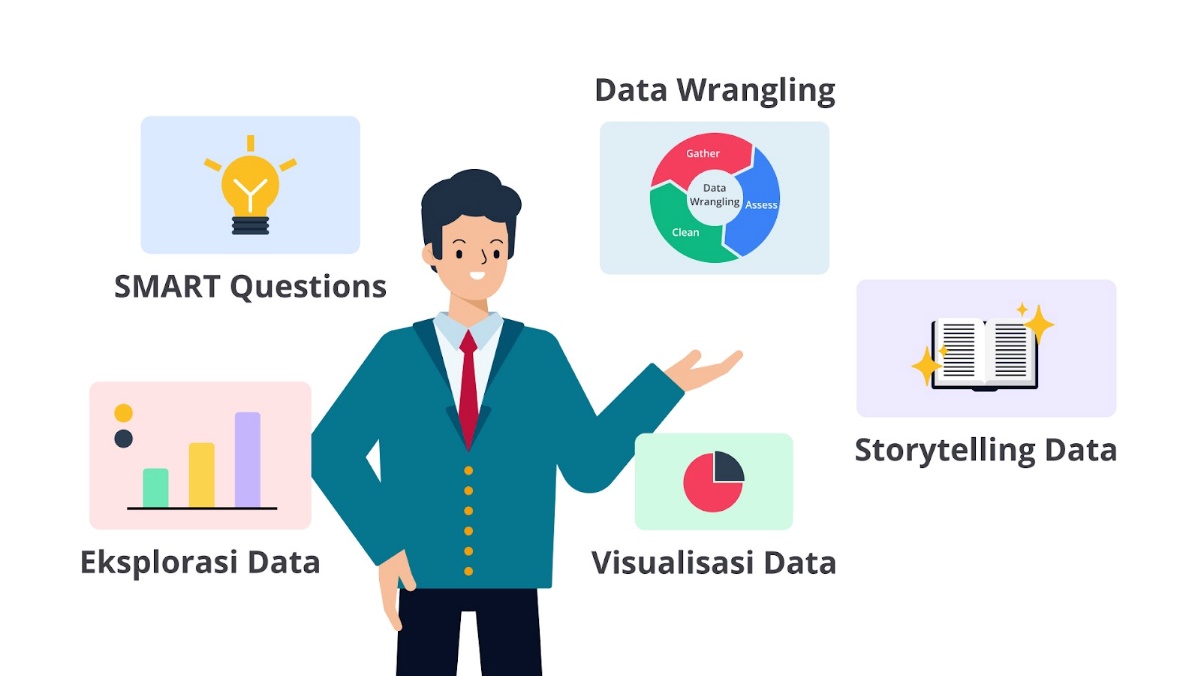
Analisis data adalah proses menyelidiki, membersihkan, mengubah, dan memodelkan data dengan tujuan menemukan informasi yang berguna, menginformasikan kesimpulan, mendukung pengambilan keputusan, dan masih banyak lagi. Untuk mewujudkan pernyataan tersebut, kita perlu melewati beberapa proses seperti yang tertera pada gambar di atas. Simak detailnya dalam uraian berikut ini.

1. **Mendefinisikan Pertanyaan**  
   Pada awal proyek analisis data, kita perlu mendefinisikan berbagai pertanyaan analisis terkait data ataupun permasalahan yang ingin diselesaikan. Pertanyaan tersebut akan membantu kita dalam memahami masalah bisnis yang dihadapi serta tujuan yang ingin dicapai. Bukan hanya itu, dengan adanya pertanyaan tersebut, kita akan merasa terbantu karena bisa lebih fokus pada bagian data yang relevan dengan tujuan.
2. **Data Wrangling**  
   Setelah memiliki pertanyaan analisis yang baik, tahap selanjutnya adalah data wrangling. Tahap ini dimulai dengan mengumpulkan data yang dibutuhkan untuk menjawab pertanyaan analisis yang telah dibuat sebelumnya. Kemudian Anda perlu menilai kualitas dan struktur dari data tersebut. Hal ini dilakukan untuk mengidentifikasi masalah yang terdapat dalam data serta membuat strategi pembersihan data yang sesuai.
3. **Exploratory Data Analysis (EDA)**  
   Tahap ini merupakan tahap eksplorasi terhadap data yang telah dibersihkan untuk memperoleh *meaningful insights*dan menjawab pertanyaan analisis. EDA memungkinkan analyst memahami isi data yang digunakan, mulai dari distribusi, frekuensi, korelasi, dan lainnya. Dalam praktiknya, memiliki rasa *curiosity*merupakan hal yang sangat penting. Pemahaman dalam konteks data juga perlu diperhatikan karena hal tersebut dapat menjawab masalah-masalah dasar.  
     
   Pada umumnya EDA dilakukan dengan beberapa cara sebagai berikut.
   * Univariat Analysis: Analisis deskriptif dengan satu variabel.
   * Bivariat Analysis: Analisis relasi dengan dua variabel yang biasanya dengan target variabel.
   * Multivariat Analysis: Analisis yang menggunakan lebih dari atau sama dengan tiga variabel.
4. **Data Visualization**  
   Hasil yang diperoleh dari EDA perlu divisualisasikan melalui tahap data visualization. Pada tahap ini, kita akan menerapkan berbagai teknik dan jenis visualisasi data yang tepat untuk mengomunikasikan temuan atau *insight*dari hasil analisis secara efektif.
5. **Draw conclusion & communicate**  
   Pada akhir proyek analisis data, kita perlu membuat *conclusion*atau kesimpulan dari hasil analisis. Kesimpulan yang dibuat haruslah menjawab semua pertanyaan dan tujuan yang telah didefinisikan di awal.

Seperti yang telah disebutkan di awal, tahapan tersebut tidak selamanya berbentuk *waterfall*atau linier. Terkadang setelah melakukan visualisasi data, Anda akan menemukan kekurangan dalam data sehingga perlu melakukan data wrangling kembali. Hal inilah yang membuat tahapan dalam analisis data lebih cocok disebut sebagai sebuah siklus karena terdapat banyak iterasi proses di dalamnya.



**Analisis Data 101**



**Studi Kasus**

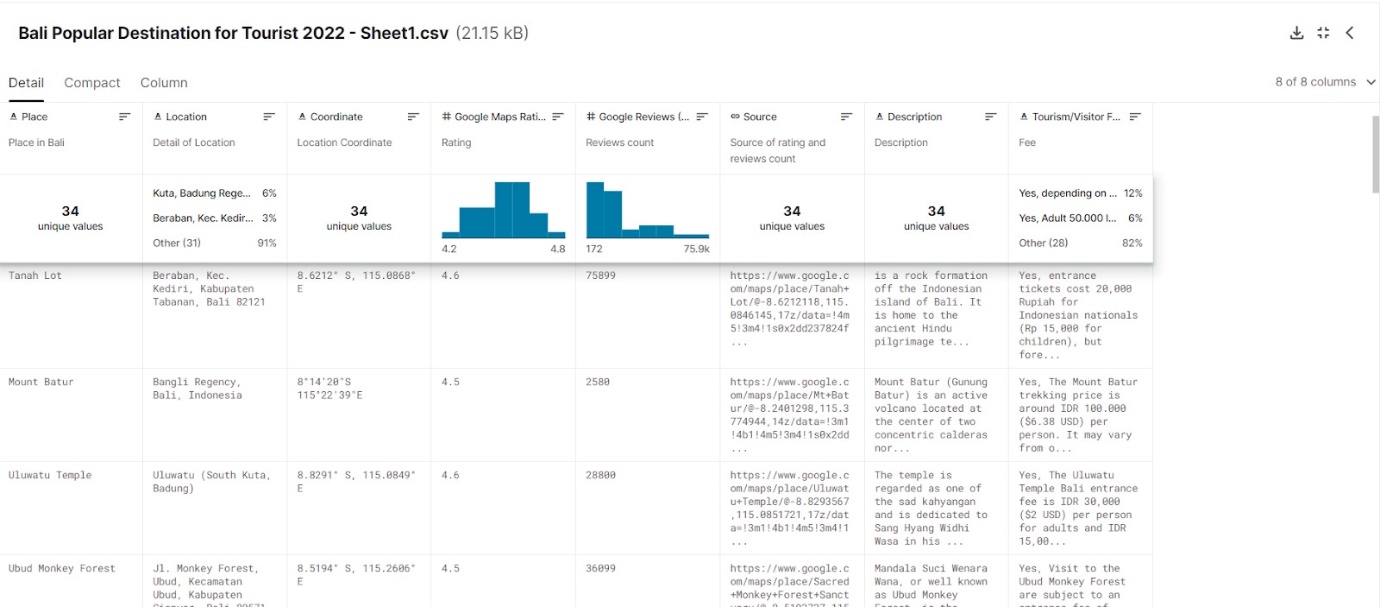
Tujuan dari studi kasus ini adalah sebagai wadah untuk latihan dalam menghadapi tantangan atau permasalahan yang terjadi di dunia nyata. Dengan studi kasus ini, diharapkan Anda lebih memahami setiap proses dalam analisis data.

Anda merupakan seorang *influencer* yang memiliki banyak pengikut di media sosial. Anda memiliki rencana untuk pergi ke Bali dengan tujuan ingin meramaikan tempat wisata di sana yang memiliki *rating* paling rendah berdasarkan *Google Maps Rating.*Sebelum berangkat ke sana, Anda mencari tahu terlebih dahulu data yang perlu dianalisis untuk dianalisis di kemudian hari agar mendapatkan hasil yang maksimal nantinya.

Berikut merupakan data yang akan Anda gunakan dalam studi kasus ini.

[Dataset Bali Popular Destination for Tourist 2022](https://www.kaggle.com/datasets/fuarresvij/bali-popular-destination-for-tourist-2022)

Ketika Anda membuka dataset tersebut, tampilannya akan terlihat seperti di bawah ini.



Nah, studi kasus ini akan kita bawa ke pembahasan analisis data yang setiap langkah-langkahnya akan dijelaskan di sana. Sudah tidak sabar? Yuk, kita lanjutkan.

**Menyusun Pertanyaan Efektif**

Untuk dapat menjelajahi sebuah data, tentunya Anda perlu memiliki *insight*atau wawasan yang luas terkait data maupun proyek yang akan dikerjakan. Tanpa sebuah informasi yang jelas, Anda tidak akan tahu ke mana arah dan tujuan dari hal yang yang ingin dikerjakan. Seperti peribahasa, “*Malu bertanya sesat di jalan*”, apabila tidak menanyakan atau mencari *insight* terkait data yang ingin dikerjakan, bisa jadi Anda akan tersesat atau kebingungan ketika mengerjakannya.

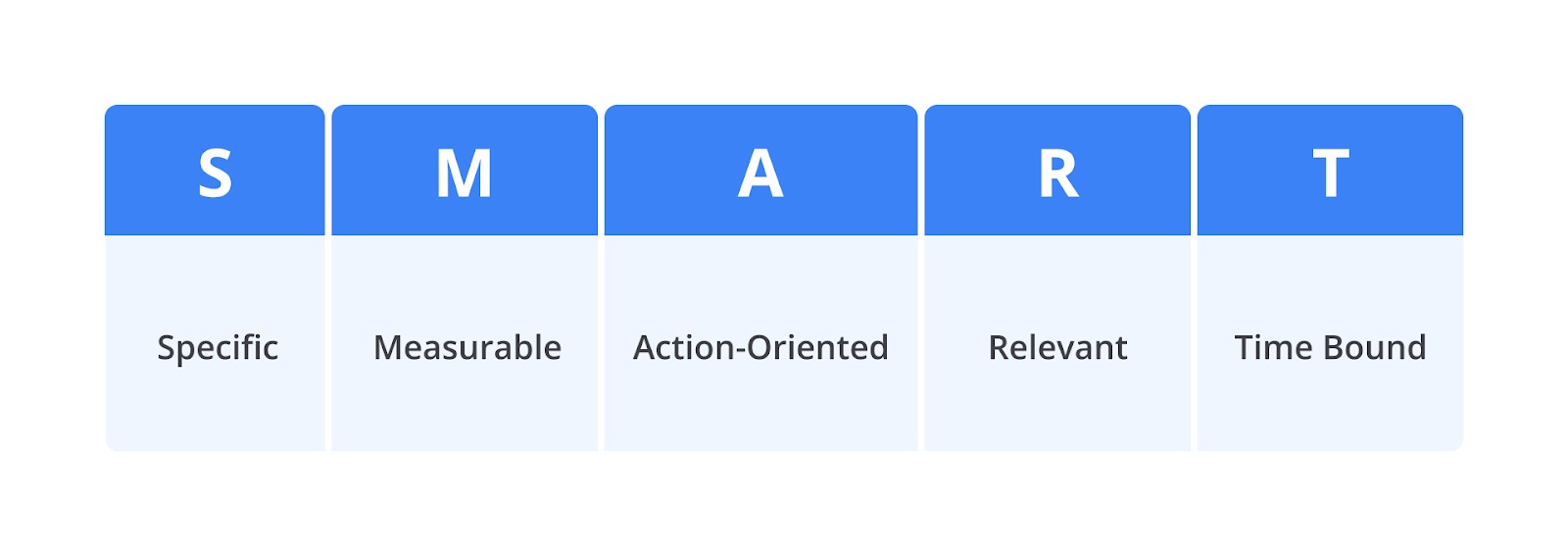
Menjadi seorang praktisi dalam menganalisis data, diibaratkan seperti Anda mendaki gunung sambil membawa sebuah kompas. Kompas merupakan sebuah perangkat yang mampu membantu Anda menentukan arah sehingga bisa sampai ke tempat tujuan. Hal tersebut juga sejalan ketika kita menyusun pertanyaan.Kita tak boleh sembarangan dan harus terarah dalam melakukannya untuk mendapatkan sebuah tujuan, yaitu *meaningful insights.*

Berikut beberapa hal yang perlu Anda hindari dalam menulis pertanyaan supaya mendapatkan hasil yang sesuai tujuan.

* Hindari membuat pertanyaan yang mengarahkan responden atau narasumber untuk menjawab dengan jawaban tertentu.  
  Contoh: Gunung Rinjani adalah gunung terindah di Indonesia, bukan?  
  Pertanyaan tersebut dihindari karena langsung merujuk kepada satu objek tertentu sehingga responden tidak dapat menyampaikan pendapat lain.
* Hindari membuat pertanyaan yang hanya dapat dijawab dengan “ya” atau “tidak”. Contoh: Apakah Anda menyukai Gunung Bromo?  
  Hal tersebut perlu dihindari karena sebuah pertanyaan diusahakan agar responden dapat menyampaikan aspirasi atau opini mereka terhadap sebuah pertanyaan.
* Hindari membuat pertanyaan yang tidak memiliki konteks.  
  Contoh**:** Apakah Pulau Komodo mahal?  
  Maksudnya adalah hal apa yang menjadi mahal di Pulau Komodo? Apakah tiket pesawat? Sewa hotel? Makanan? Konteks pada pertanyaan di atas belum terlalu jelas sehingga perlu dihindari dalam membuat sebuah pertanyaan.

Setelah mengetahui beberapa hal yang harus dihindari dalam menyusun pertanyaan, ada sebuah metode yang perlu Anda ketahui dan terapkan supaya dapat menghasilkan pertanyaan analisis yang efektif. Metode apakah itu?

**SMART Questions**



Mari kita *breakdown*satu per satu kepanjangan dari SMART.

* **Specific**  
  Pertanyaan efektif tentu harus bersifat spesifik, artinya sebuah pertanyaan analisis harus merujuk pada topik tertentu. Pertanyaan spesifik mampu membantu Anda dalam mendapatkan jawaban yang spesifik. Oleh karena itu, apabila sebuah pertanyaan masih bersifat general, cobalah untuk lebih mengerucut lagi pada satu topik tertentu. Tentunya dengan memperhatikan hal yang harus dihindari seperti yang sudah dijelaskan sebelumnya.
* **Measurable**  
  Selain spesifik, hal yang perlu diperhatikan adalah sebuah pertanyaan mesti terukur (*measurable*). Jawaban dari hal ini akan menjadi sebuah tolok ukur Anda dalam mendapatkan jawaban yang *meaningful insights.*
* **Action-oriented**  
  Maksud dari Action-oriented adalah pertanyaan yang diajukan harus mampu memberikan informasi yang dapat membantu Anda dalam membuat sebuah rencana yang harus dilakukan nantinya*.*
* **Relevant**  
  Poin keempat ini memiliki arti bahwa pertanyaan yang diajukan harus berbobot dan memiliki arti untuk menyelesaikan sebuah permasalahan yang ingin diselesaikan.
* **Time-bound**  
  Poin terakhir yaitu pertanyaan yang diajukan mesti bersifat *time-bound* atau selaras terhadap waktu saat ini.

Buat pertanyaan menggunakan SMART terkait studi kasus dari dataset “**Bali Popular Destination for Tourist 2022**” yang telah diberikan pada materi sebelumnya untuk digunakan pada langkah-langkah berikutnya.

Contoh:

* **Specific**: “Destinasi wisata manakah yang memiliki rating paling rendah?”
* **Measurement**: “Berapa banyak destinasi wisata yang memiliki rating di bawah rata-rata?” dan “Berapa rating terendah yang terdapat pada destinasi wisata tersebut?”
* **Action Oriented**: “Hal apa yang perlu dilakukan daerah setempat untuk meningkatkan *rating*destinasi wisata tersebut?”
* **Relevant**: “Apakah aksi tersebut dapat meningkatkan *rating* destinasi wisata? ”
* **Time-Bound**: “Berapa lama rencana peningkatan tersebut dilaksanakan?”

**Data Wrangling**

Data wrangling merupakan sebuah proses atau kumpulan kegiatan yang meliputi mengumpulkan data (*gathering data*), menyiapkan data untuk dianalisis (*assessing data*), dan pembersihan data (*cleaning data*) sebelum data digunakan dalam proses analisis.

Untuk dapat melalui data wrangling, Anda akan melewati beberapa langkah yang merupakan lanjutan dari langkah 1 sebelumnya. Mari kita mulai dengan mengumpulkan data.

*Disclaimer*: dalam kondisi nyata di lapangan, bisa saja data-data yang telah dikumpulkan masih terdapat kekurangan sehingga Anda harus mencari sumber relevan untuk memperluas data, seperti hasil dari wawancara warga setempat, kuesioner wisatawan, jurnal, website atau sumber lainnya. Data-data tersebut nantinya dikumpulkan terlebih dahulu agar semuanya sudah siap ketika Anda memerlukannya untuk menjawab semua pertanyaan atau masalah bisnis yang akan dihadapi.

**Menyiapkan Data untuk Dianalisis**

Setelah melewati proses mengumpulkan data, pada tahap ini bisa juga disebut sebagai *assessing data*dengan kata lain penilaian terhadap data tersebut.

**Langkah 3**

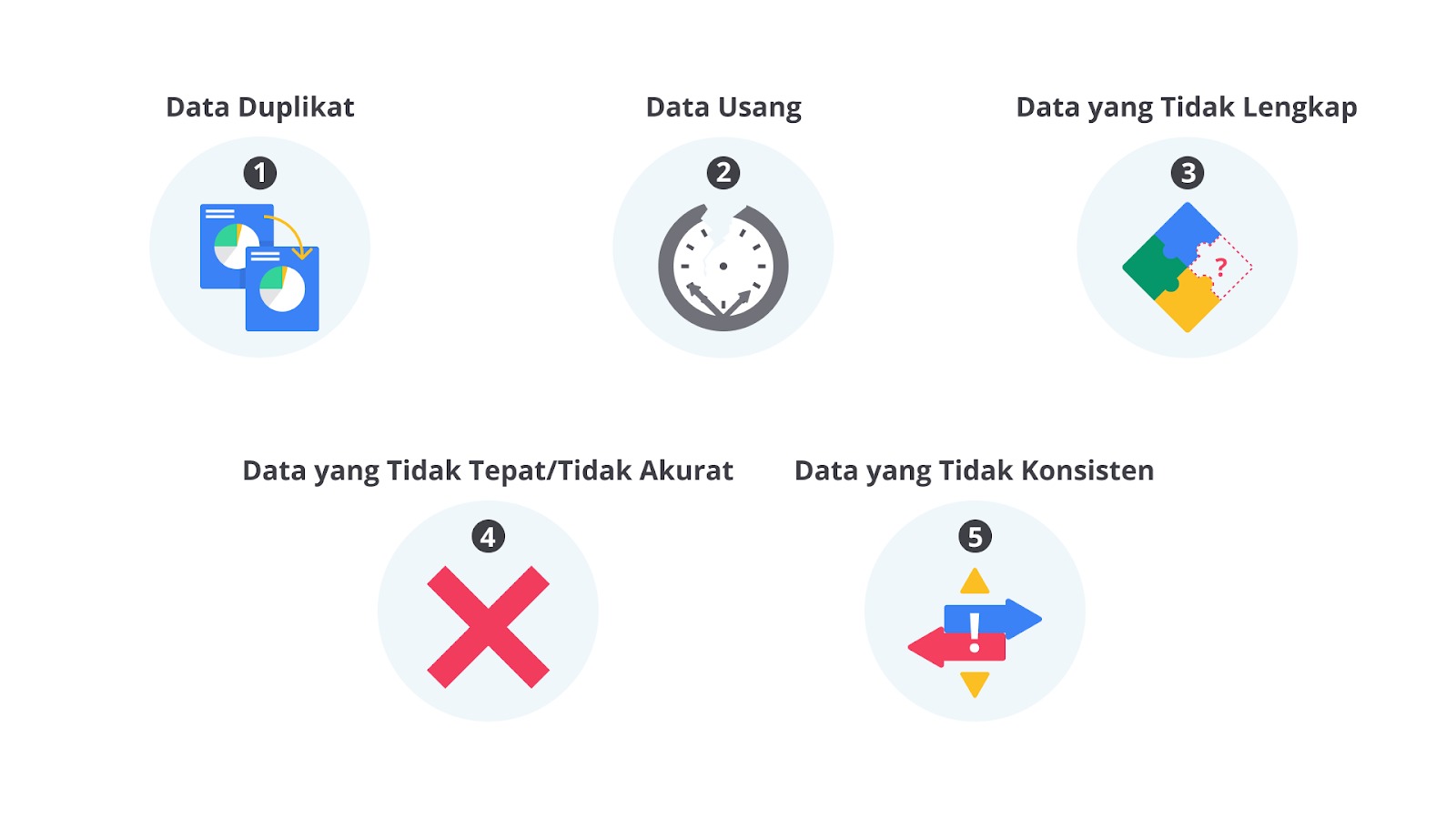
Dalam proses data wrangling, setelah Anda mengumpulkan data, hal selanjutnya yang perlu dilakukan adalah menyiapkan data. Langkah ini dilakukan untuk menilai kualitas dan struktur dari sebuah data. Selain itu, proses ini juga bertujuan untuk mengidentifikasi berbagai masalah yang terdapat dalam data, seperti *missing value, understand value,*dll.

**Mari Kita Bersihkan!**

Sebuah data yang telah terkumpul tentu tidak semena-mena dapat langsung digunakan karena memerlukan proses panjang sebelum ia dapat menghasilkan data yang informatif.

Sebelum membersihkan data-data yang telah didapatkan, apakah Anda telah paham kriteria yang termasuk dalam data kotor? Jika belum, simak penjelasannya di bawah ini.

Data kotor adalah data yang tidak lengkap, salah, atau tidak relevan dengan masalah yang ingin diselesaikan. Berikut merupakan tipe-tipe data kotor.



**Duplicate data**

Duplicate data adalah data apa pun yang muncul lebih dari satu kali. Hal tersebut bisa jadi diakibatkan oleh entri data, impor, atau migrasi data secara manual. Apabila tetap dijalankan, perhitungan atau prediksi tidak akan akurat sehingga akan mempengaruhi proses analisis selama pengambilan data.

**Outdated data**  
Setiap data yang sudah tua harus diganti dengan informasi yang lebih baru dan akurat. Hal tersebut disebabkan oleh peran atau perusahaan yang berganti atau bisa jadi *software*dan sistem yang telah usang. Apabila tetap dijalankan, beberapa hal seperti wawasan, pengambilan keputusan, dan analisis tidak akan akurat. *Outdate data*biasanya berupa data yang usang atau kadaluwarsa sehingga bisa jadi sudah tidak relevan dengan kondisi saat ini. Misalnya, Anda menggunakan data destinasi wisata tahun 2013. Apabila dibandingkan dengan kondisi saat ini, mungkin akan ada perubahan karena dalam satu dekade bisa saja terjadi bencana alam, pandemi, dan lain-lain.

**Incomplete data**

Incomplete data merupakan kondisi ketika data apa pun yang tidak memiliki bidang penting atau tidak komplit. Hal tersebut disebabkan oleh pengumpulan data yang tidak benar atau entri data yang salah. Apabila tetap dijalankan, akan terjadi penurunan produktivitas, wawasan yang tidak akurat, atau ketidakmampuan untuk menyelesaikan layanan penting.

**Incorrect/inaccurate data**

Jumlah data yang banyak bisa jadi data tersebut adalah tidak akurat. Ketidakakuratan data bisa disebabkan oleh kesalahan manusia *(human error*), informasi palsu, atau data tiruan. Apabila hal tersebut tetap dijalankan, wawasan atau pengambilan keputusan yang tidak akurat berdasarkan informasi buruk bisa jadi akan mengakibatkan hilangnya pendapatan.

**Inconsistent data**

Data apa pun yang menggunakan format berbeda untuk mewakili hal yang sama akan menyebabkan ketidakkonsistenan. Hal tersebut bisa disebabkan oleh data yang tersimpan secara tidak benar atau terjadi kesalahan penyisipan selama transfer data. Apabila hal tersebut tetap dijalankan, poin data yang kontradiktif akan menyebabkan kebingungan atau ketidakmampuan untuk mengklasifikasikan atau mengelompokkan pelanggan.

**Mengapa Membersihkan Data Begitu Penting?**

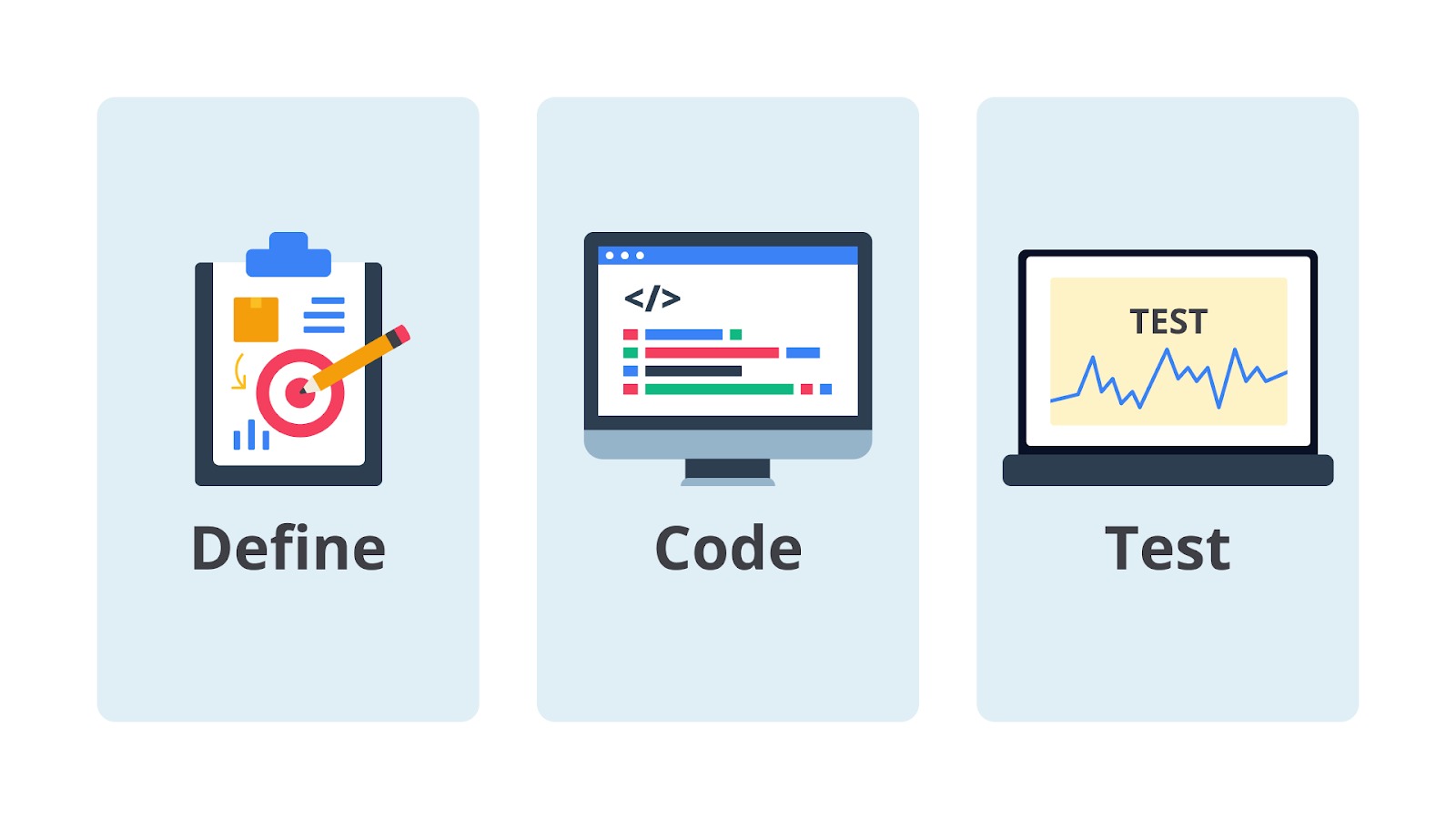
Tahukah Anda bahwa pembersihan data yang tidak akurat atau buruk dapat merugikan bisnis setiap tahun? Apakah ribuan dollar? Jutaan? Miliaran? Nah, menurut IBM, biaya tahunan untuk membuat data yang berkualitas membutuhkan $3.1 triliun untuk di AS saja.

Hal tersebut bukan karena kesalahan sistem atau teknis komputer, melainkan kesalahan manusia yang mengakibatkan adanya data kotor atau data yang tidak relevan pada setiap proses analisis. Ketika bekerja dengan data kotor, Anda tidak dapat memastikan bahwa hasil semua data adalah benar. Integritas data sangat penting untuk hasil analitik data yang andal. Selain itu, data yang bersih juga akan membantu Anda untuk mencapai integritas data.

Jadi, alasan dari pentingnya proses pembersihan data itu sangat sederhana yaitu agar dapat menghasilkan kualitas data pada analisis.

**Cara Membersihkan Data**

Rekan Anda akan memberikan sebuah daftar masalah umum yang dapat Anda rujuk saat melakukan verifikasi pembersihan data.



Secara umum, proses pembersihan data dapat dibagi ke dalam tiga tahapan, yaitu *define, code,*dan *test.*

* **Define.** Pada tahap ini, kita akan membuat rancangan tahapan serta metode pembersihan data berdasarkan masalah yang ditemukan dalam proses assessing data. Hal ini dapat dijadikan sebagai dokumentasi untuk memastikan orang lain memahami setiap tahapan dalam pembersihan data yang akan kita lakukan.
* **Code.** Setelah membuat rancangan pembersihan data, tahap selanjutnya ialah mengonversi hal tersebut menjadi sebuah kode program yang dapat dijalankan.
* **Test.** Setelah menjalankan kode program untuk membersihkan data, kita perlu memeriksa kembali data yang telah dibersihkan tersebut. Hal ini untuk memastikan proses pembersihan data dilakukan sesuai ekspektasi kita.

Mengeksplorasi Data (Exploratory Data Analysis)

Pada dasarnya tujuan eksplorasi data adalah untuk mengidentifikasi pola, hubungan (korelasi) dan tren dalam data sehingga Anda dapat menjawab pertanyaan yang Anda ajukan secara akurat.

Sebelum melakukan eksplorasi data, hal lain yang perlu Anda ketahui bahwa korelasi memiliki nilai dalam rentang tertentu untuk menentukan hubungan yang makin kuat, yaitu -1 (negatif) dan 1 (positif). Dengan diketahuinya nilai korelasi maka dapat melihat seberapa kuat korelasi antar data tersebut. Setelah mengetahui nilai korelasi tersebut dapat dilihat tren dari data tersebut apakah cukup kuat atau tidak.

Namun, apakah sebelumnya Anda telah mengetahui definisi dari tren yang dimaksud?

**Tren** merupakan sebuah analisis dengan membandingkan data dari waktu ke waktu untuk melihat hasil yang konsisten.

Maka dari itu, hasil akhir dari eksplorasi data ini akan menjadi sebuah jawaban untuk SMART Questions yang sebelumnya telah Anda rumuskan.

Nah, di bawah ini merupakan empat fase eksplorasi secara umum untuk mendapatkan tren dari sebuah data.

1. **Mengatur data**  
   Memastikan data sudah siap digunakan, bukan lagi data “mentah” yang harus diolah lagi.
2. **Memformat dan menyesuaikan data**  
   Sesuaikan data yang ingin dilihat korelasi satu sama lain untuk memudahkan analisis.
3. **Mendapatkan *insights*dari orang lain**  
   Selain menganalisis berdasarkan data, Anda juga perlu mengeksplorasi dari berbagai cara untuk mendapatkan *insights.*Contohnya melalui wawancara warga setempat terkait faktor yang memengaruhi destinasi wisata tersebut mendapatkan rating rendah atau mencari tahu melalui website destinasi wisata tersebut, dll.
4. **Amati hubungan antar titik data dan membuat perhitungan**

**Memvisualisasi Data**

Setelah sekian *purnama* Anda bergelut dengan data, pada materi ini kita masuk ke pembahasan visualisasi data. Dari seluruh tahap menyusun pertanyaan efektif, data wrangling, hingga mengeksplorasi data, seharusnya Anda telah memiliki sebuah data yang sudah siap untuk digunakan.

Nah, data-data tersebut nantinya akan diinterpretasikan ke dalam grafik dalam tahap memvisualisasikan data ini.



Hal ini dilakukan untuk mempermudah kita dalam memahami data dan membuat sebuah kesimpulan yang tepat. Selain itu, visualisasi data juga dapat membantu kita dalam menyampaikan cerita atau temuan dari hasil analisis data kepada orang lain atau stakeholder.

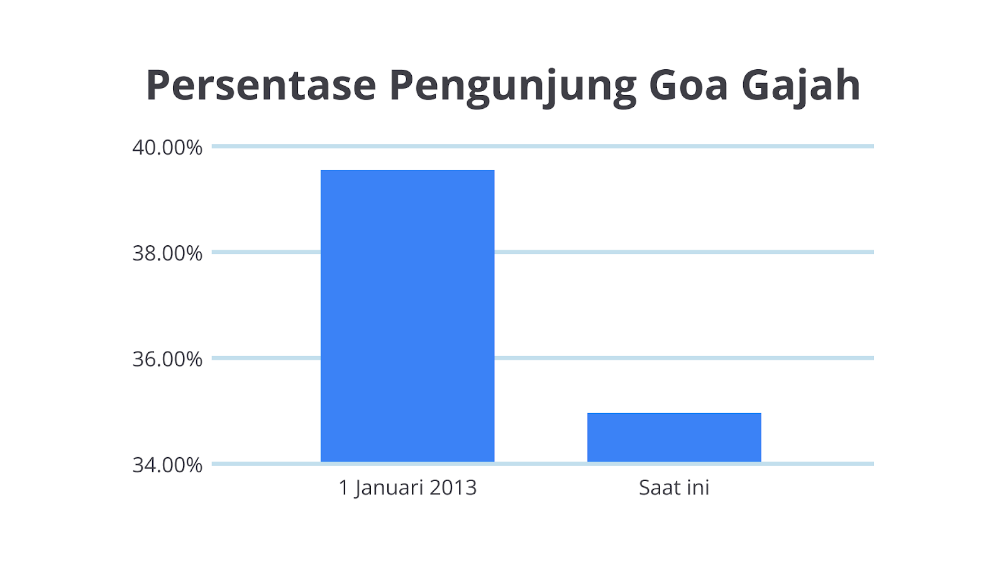
**Jenis-jenis Visualisasi Data yang Dihindari**

Seperti yang telah disampaikan di awal, pada beberapa materi ke depan, kita akan fokus pada pembahasan visualisasi data. Ketika berbicara tentang visualisasi data, kita tidak hanya membahas cara menyajikan data ke dalam bentuk visual, melainkan juga harus memahami berbagai prinsip dasar dalam membuat visualisasi data yang baik dan benar.

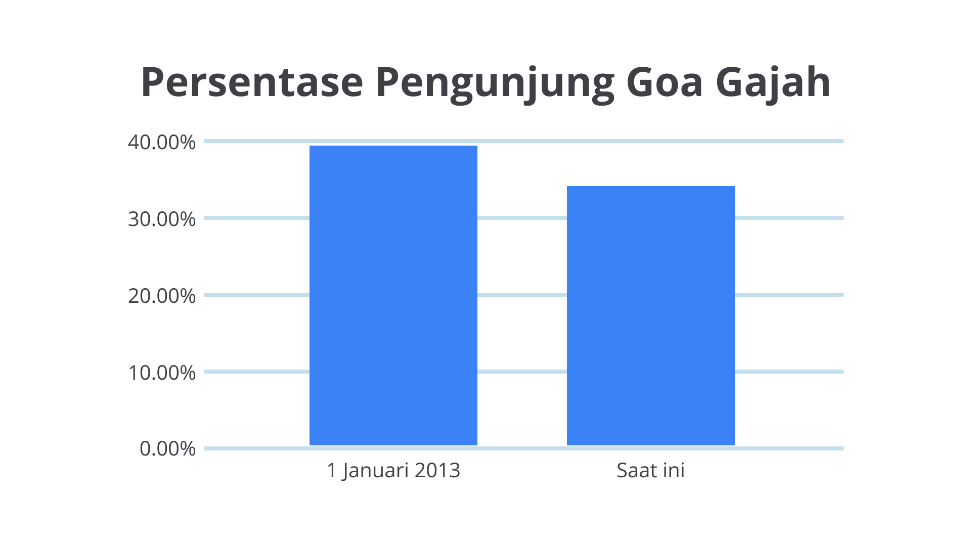
Nah, pada materi ini, kita akan fokus membahas tentang prinsip tersebut. Sebagai permulaan, mari kita lihat berbagai contoh visualisasi data yang buruk dan tidak mengikuti prinsip desain dalam visualisasi data.

Sederhananya, visualisasi data yang buruk merupakan bentuk visual yang tidak mampu menyampaikan pesan dan informasi terkait data secara baik serta efisien. Umumnya visualisasi data yang buruk memuat salah satu dari tiga masalah berikut.

* **Misleading**  
  Masalah pertama dalam visualisasi data yang buruk ialah *misleading information* yang mampu mengakibatkan kesalahan dalam pengambilan kesimpulan dari sebuah data. Berikut contoh visualisasi data yang bisa mengakibatkan *misleading information*.

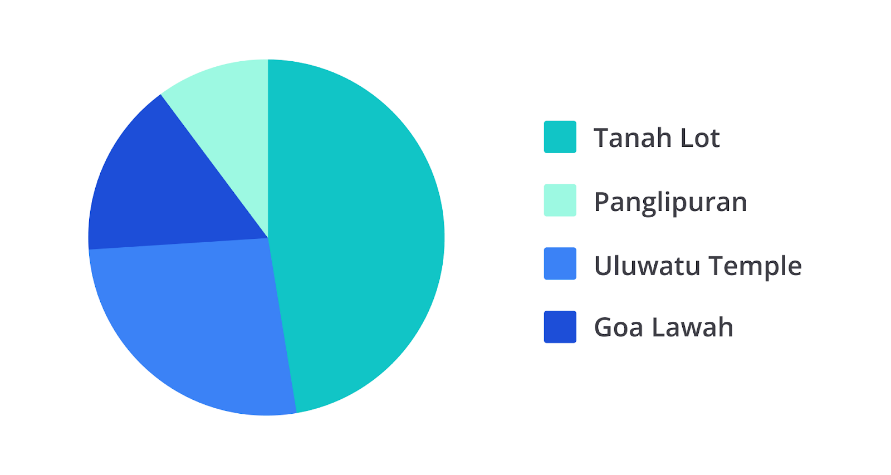


Apa yang ada di benak Anda ketika melihat visualisasi data tersebut? Tentunya Anda akan melihat bahwa terdapat gap yang cukup besar dari kedua nilai tersebut. Hal ini contoh kesalahan pengambilan kesimpulan karena buruknya sebuah visualisasi data.  
  
Apakah Anda bisa menebak kesalahan apa yang mengakibatkan visualisasi data di atas menghasilkan kesimpulan yang salah? *Yap,*benar sekali visualisasi datanya tidak mulai dari nol sehingga terkesan ada gap yang cukup besar dari kedua nilai tersebut. Berikut merupakan contoh perbaikan dari visualisasi di atas.

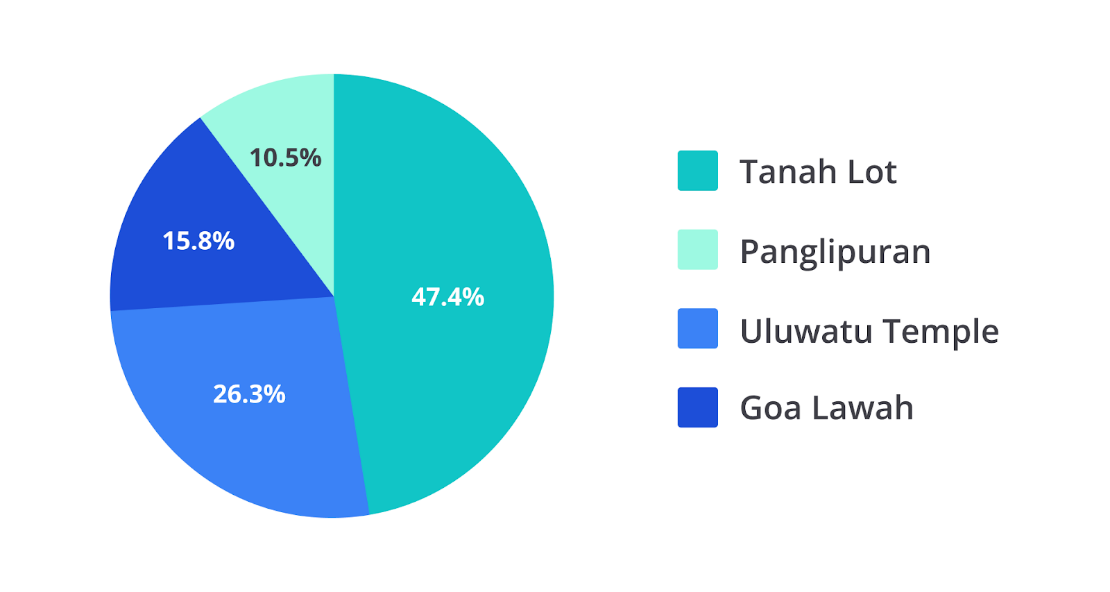


Anda dapat melihat dari grafik di atas bahwa ternyata tidak ada *gap*yang terlalu besar antara dua data tersebut. Itulah pentingnya membuat grafik dari angka 0.

* **Hides**  
  Selain *misleading information*, masalah lain yang umum dijumpai dalam visualisasi data yang buruk ialah menyembunyikan informasi tertentu. Berikut contoh visualisasi data yang menyembunyikan informasi penting dari sebuah data.



Berdasarkan visualisasi di atas, bisakah Anda menjawab destinasi wisata yang paling banyak dikunjungi? Anda mungkin akan kesulitan dalam menjawab kedua pertanyaan tersebut karena banyak sekali informasi yang disembunyikan.  
  
Maka dari itu perlu ditampilkan dengan jelas persentase dari nilai-nilai grafik di atas seperti yang tertera di bawah ini.



Apabila dibandingkan dengan gambar sebelumnya, tentu pada gambar ini Anda dapat menjawab pertanyaan destinasi wisata yang paling banyak dikunjungi, bukan?

* **Distract**  
  Masalah lain yang mungkin terjadi ialah distraksi. Visualisasi data yang buruk sering kali menyertakan komponen visual yang seharusnya tidak dibutuhkan sehingga mengganggu proses pengambilan kesimpulan. Berikut merupakan contoh visualisasi data yang memuat komponen visual yang tidak dibutuhkan.



Pada contoh di atas, tidak terdapat *highlight*warna pada data untuk menunjukkan nilai tertinggi atau terendah.  
  
Warna merupakan komponen visual yang penting dalam visualisasi data. Oleh karena itu, kita bisa menggunakannya untuk menyampaikan pesan yang menarik dan mempermudah orang lain untuk memahami data yang kita miliki. Gambar di bawah ini merupakan contoh visualisasi data yang baik dalam penggunaan warna untuk menyorot bagian tertentu.



* Pada gambar tersebut terdapat warna yang berbeda yaitu pada bagian “Bali Zoo”, warna tersebut di *highlight*karena memiliki nilai yang paling tinggi di antara lainnya.

Oke, kita telah melihat berbagai masalah beserta contohnya pada visualisasi data yang buruk. Kini saatnya kita mengenal berbagai prinsip dalam visualisasi data untuk menghindari kesalahan tersebut.

**Prinsip-Prinsip dalam Memvisualisasikan Data**

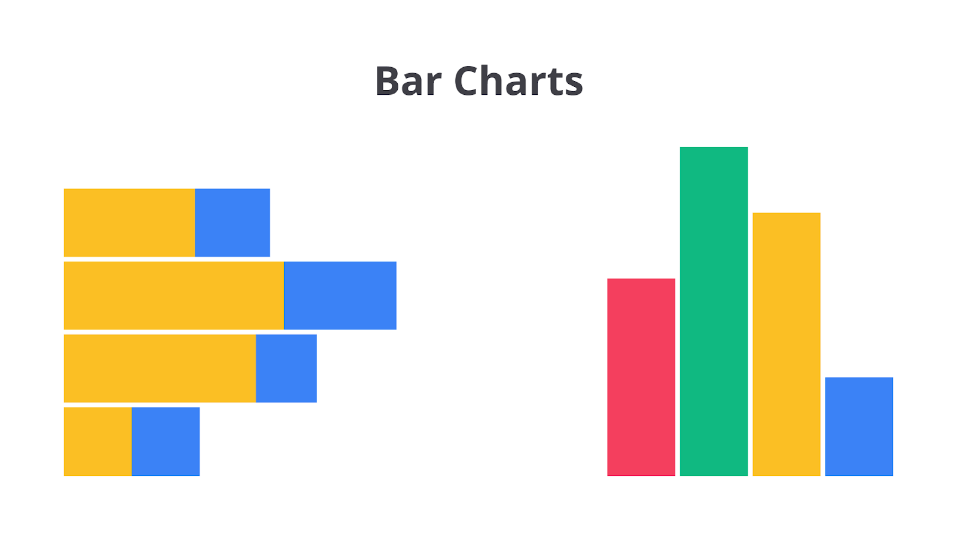
Jika ditelisik secara mendasar, visualisasi data merupakan proses dalam mengubah data ke dalam bentuk visual menggunakan berbagai elemen visual. Berikut merupakan beberapa prinsip dalam membangun elemen visual yang umum digunakan untuk membuat visualisasi data.

* **Position**. Elemen ini akan membantu kita merepresentasikan titik data menggunakan sumbu tertentu (seperti sumbu X, Y, dan Z) sebagai acuan.
* **Size**. Ukuran (panjang atau lebar) merupakan elemen visual yang umumnya kita gunakan untuk membedakan serta membandingkan nilai dari kategori atau titik data tertentu.
* **Shape**. Bentuk merupakan salah satu elemen visual yang dapat digunakan untuk membedakan kategori atau titik data tertentu.
* **Color**. Selain bentuk, warna juga merupakan pilihan elemen visual lain yang dapat digunakan untuk membedakan kategori atau titik data tertentu. Ketika menggunakan elemen ini, kita harus ingat bahwa tidak semua orang memiliki kemampuan untuk membedakan warna dengan baik.
* **Texture**. Penambahan tekstur atau pola tertentu bisa menjadi alternatif lain dalam membedakan kategori atau titik data tertentu.
* **Angle**. Pada beberapa pilihan bentuk visualisasi data, sudut merupakan salah satu elemen visual yang digunakan untuk merepresentasikan nilai dari suatu data.

**Eksplorasi Grafik**

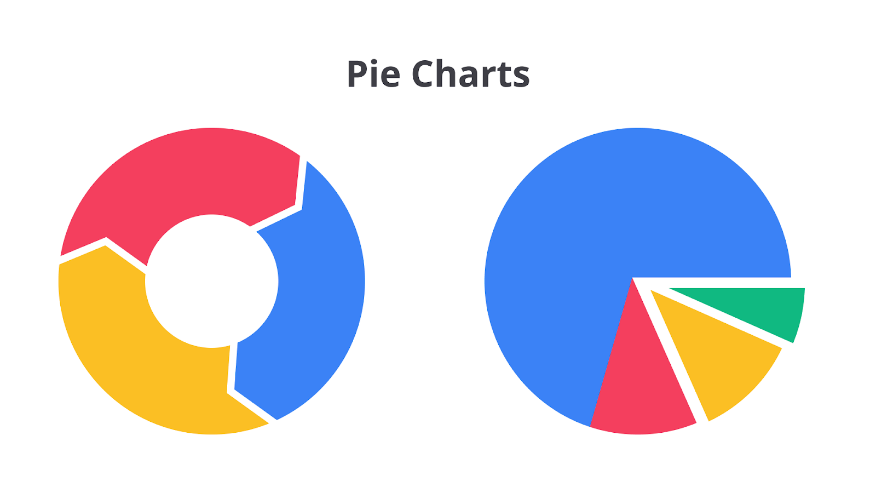
Tahukah Anda? Ada lebih dari 40 jenis grafik di luar sana, tetapi lebih kurang hanya beberapa yang lebih umum untuk digunakan. Mari kita eksplor tentang 5 jenis bagan teratas yang digunakan dan waktu untuk menggunakannya.

**Bar Chart**



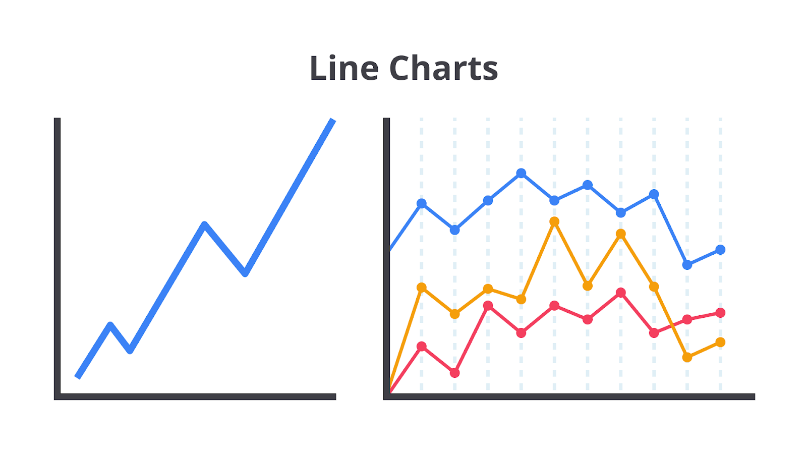
* Kapan menggunakannya?
  + Ketika membandingkan bagian dari sekumpulan data yang lebih besar, menyoroti kategori yang berbeda, atau menampilkan perubahan dari waktu ke waktu.
  + Ketika ingin mengilustrasikan nilai positif dan negatif dalam kumpulan data.
* Kapan harus dihindari?
  + Ketika Anda memiliki banyak kategori karena grafik batang sebaiknya tidak boleh memiliki lebih dari 10 batang (kategori).

**Pie Chart**



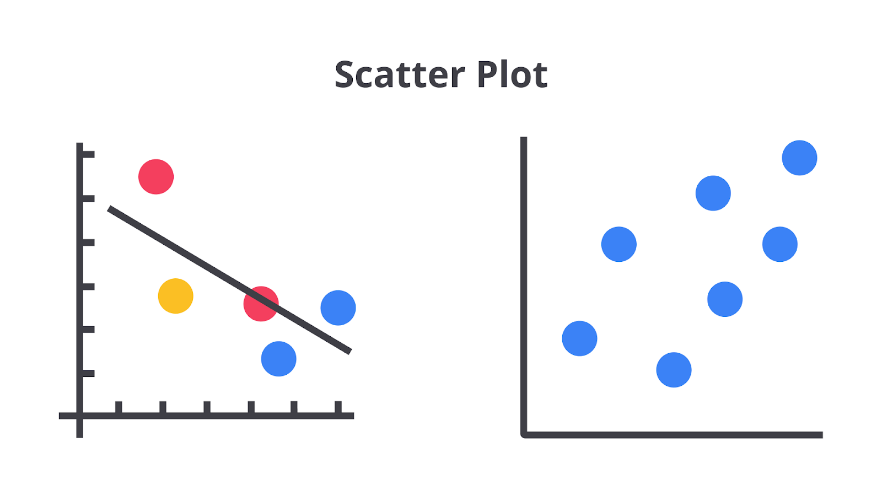
* Kapan menggunakannya?
  + Saat Anda menampilkan proporsi dan persentase relatif dari seluruh kumpulan data.
  + Paling baik digunakan dengan kumpulan data kecil.
  + Ketika data yang Anda memiliki tidak lebih dari 6 kategori.
  + Ketika data Anda nominal dan bukan ordinal.
* Kapan harus dihindari?
  + Ketika Anda memiliki kumpulan data yang besar.
  + Ketika Anda ingin membuat perbandingan yang tepat atau absolut antara nilai.

**Line Chart**



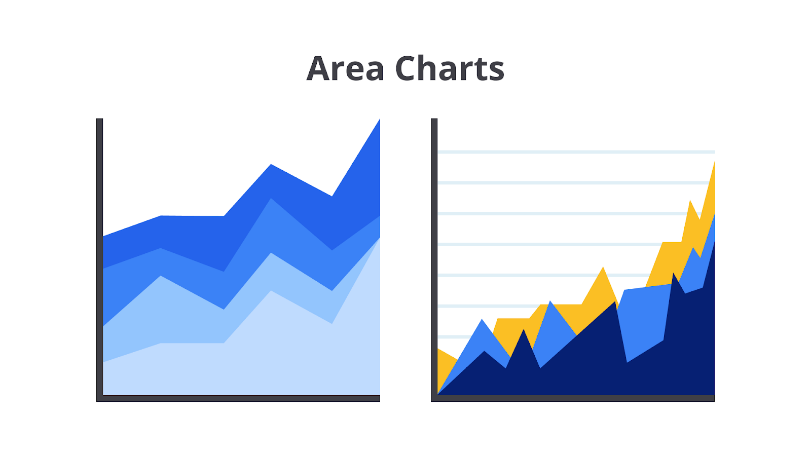
* Kapan menggunakannya?
  + Ketika memiliki kumpulan data berkelanjutan yang berubah seiring waktu.
  + Saat kumpulan data terlalu besar untuk bar chart.
  + Ketika ingin menampilkan beberapa seri untuk garis waktu yang sama.
  + Saat ingin memvisualisasikan tren, bukan nilai yang tepat.
* Kapan harus dihindari?
  + Line chart bekerja lebih baik dengan kumpulan data yang lebih besar, jadi ketika hanya terdapat data yang kecil atau sedikit, maka bisa gunakan bagan batang.

**Scatter plot**



* Kapan menggunakannya?
  + Untuk menunjukkan korelasi dan pengelompokkan dalam kumpulan data besar.
  + Ketika kumpulan data berisi titik yang memiliki sepasang nilai.
  + Ketika urutan poin dalam dataset tidak penting.
* Kapan harus dihindari?
  + Ketika terdapat kumpulan data kecil.
  + Ketika nilai dalam kumpulan data tidak berkorelasi satu sama lain.

**Area Chart**



* Kapan menggunakannya?
  + Ketika Anda ingin menunjukan hubungan bagian keseluruhan
  + Ketika ingin menggambarkan volume data dan bukan hanya hubungannya dengan waktu
* Kapan harus dihindari?
  + Tidak dapat digunakan dengan data diskrit

**Bercerita dengan Data**

Sebelum Anda mempelajari hal ini lebih lanjut, terdapat hal yang perlu Anda ketahui bahwa fakta dan angka sangat penting dalam dunia bisnis, tetapi jarang memberikan kesan yang bertahan lama dalam ingatan para pendengarnya. Berjalan dari latar belakang tersebut untuk menciptakan komunikasi yang membuat orang berpikir dan meyakinkan mereka untuk mengambil tindakan, maka diperlukan sebuah penceritaan data. Nah, pada pembahasan kali ini Anda akan belajar terkait cara menceritakan sebuah data kepada para *audiens.*Sudah tidak sabar? Yuk, simak penjelasannya di bawah ini.

Setelah Anda melalui semua tahap dalam menganalisis data, berikut merupakan langkah terakhir yang perlu Anda lakukan guna semua tahap terpenuhi dan Anda mampu menjadi seorang data scientist yang Andal, yaitu bercerita dengan data atau biasa disebut *storytelling.*

*Storytelling*adalah mengomunikasikan makna kumpulan data dengan visual dan narasi yang disesuaikan untuk setiap audiens tertentu. Narasi adalah kata lain dari cerita.

Berikut merupakan tiga hal yang perlu Anda siapkan sebelum bercerita dengan data.

1. **Libatkan audiens.** Ketika Anda menganalisis sebuah data, maka libatkan audiens untuk melihat hasil dari analisis tersebut. Karena hal tersebut adalah cara yang jauh lebih menarik untuk berbagi data.
2. **Buat visual yang menarik.** Anda ingin menunjukkan kisah data Anda, bukan hanya menceritakannya. Visual harus membawa audiens Anda dalam perjalanan tentang bagaimana data berubah dari waktu ke waktu.
3. **Ceritakan kisah dengan cara yang menarik**. Tahap terakhir ini adalah menceritakan kisah tersebut dalam narasi yang menarik.

Berikut merupakan contoh dari *storytelling*yang dapat Anda gunakan apabila membuat sebuah portofolio.

[Download Contoh Storytelling Data](https://docs.google.com/document/d/1GBkji8l_QTwqDAOfSBxq47eesAdSM4LazKOt1ZpuG68/copy)