**Rangkuman Menjelajahi Analisis Data**

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

**Mengumpulkan Data**

Setelah mengajukan pertanyaan efektif menggunakan SMART Questions, yang harus Anda lakukan selanjutnya adalah mengumpulkan data. Data-data yang telah didapatkan baik dari kuesioner atau wawancara atau sumber lainnya, perlu dikumpulkan terlebih dahulu tujuannya adalah ketika Anda memerlukan data-data yang dibutuhkan untuk menjawab semua pertanyaan atau masalah bisnis yang akan dihadapi sudah siap.

**Menyiapkan Data untuk Dianalisis**

Tahap ini bisa juga disebut sebagai *assessing data*dengan kata lain penilaian terhadap data tersebut. Proses 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!**

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

Mari kita *breakdown*satu persatu agar Anda semakin paham.

1. **Duplicate data**Duplicate data adalah catatan 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, maka perhitungan atau prediksi tidak akan akurat, atau akan bingung selama pengambilan data.
2. **Outdated data**  
   Setiap data yang sudah tua harus diganti dengan informasi yang lebih baru dan lebih akurat. Hal tersebut disebabkan oleh peran atau perusahaan yang berganti atau bisa jadi *software*dan sistem yang telah usang. Apabila tetap dijalankan, maka wawasan, pengambilan keputusan, dan analisis tidak akan akurat.
3. **Incomplete data**  
   Incomplete berarti tidak komplit adalah kondisi ketika data apa pun yang tidak memiliki bidang penting. Hal tersebut disebabkan oleh pengumpulan data yang tidak benar atau entri data yang salah. Apabila tetap dijalankan, maka akan penurunan produktivitas, wawasan yang tidak akurat, atau ketidakmampuan untuk menyelesaikan layanan penting.
4. **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.
5. **Inconsistent data**  
   Data apapun yang menggunakan format berbeda untuk mewakili hal yang sama. Ketidakkonsistenan data bisa saja disebabkan oleh data disimpan secara tidak benar atau kesalahan menyisipkan selama transfer data. Apabila hal tersebut tetap dijalankan, maka poin data yang kontradiktif menyebabkan kebingungan atau ketidakmampuan untuk mengklasifikasikan atau mengelompokkan pelanggan.

**Mengapa Membersihkan Data Begitu Penting?**

Fungsi pembersihan data adalah

* mengidentifikasi pola-pola penting
* menghubungkan informasi terkait, dan
* menarik kesimpulan yang bermanfaat.

Oleh karena itu, mengapa pembersihan data begitu diperlukan? Jawabannya sangat sederhana yaitu agar menghasilkan kualitas data pada analisis.

**Cara Membersihkan 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**

Tujuan analisis adalah untuk mengidentifikasi tren dan hubungan dalam data sehingga Anda dapat menjawab pertanyaan yang Anda ajukan secara akurat. Sebelum melakukan eksplorasi data, apakah yang dimaksud dengan tren?

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

Di bawah ini merupakan empat fase analisis pada umumnya, yaitu

* mengatur data.
* memformat dan menyesuaikan data.
* mendapatkan *insights*dari orang lain.
* mengubah data dengan mengamati hubungan antara titik data dan membuat perhitungan.

**Menyusun Pertanyaan Efektif**

* **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 dengan memperhatikan hal yang harus dihindari di atas, maka coba lebih mengerucut lagi pada satu topik tertentu.
* **Measurable**  
  Selain spesifik yang perlu diperhatikan adalah sebuah pertanyaan mesti terukur (*measurable*). Jawaban dari hal ini akan menjadi sebuah tolak ukur Anda dalam mendapatkan jawaban yang *meaningful insights.*
* **Action-oriented**  
  Kategori ini bermaksud bahwa pertanyaan yang diajukan harus mampu memberikan informasi yang dapat membantu Anda dalam membuat sebuah *action-plan.*
* **Relevant**  
  Poin keempat ini memiliki arti bahwa pertanyaan yang diajukan harus berbobot dan memiliki arti untuk menyelesaikan sebuah permasalahan.
* **Time-bound**  
  Poin terakhir yaitu pertanyaan yang diajukan mesti bersifat time-bound atau memiliki relevansi terhadap waktu saat ini.

**Jenis-Jenis Visualisasi Data yang Dihindari**

* **Misleading**  
  Masalah pertama dalam visualisasi data yang buruk ialah *misleading information* yang mampu mengakibatkan kesalahan dalam pengambilan kesimpulan dari sebuah data.
* **Hides**  
  Selain *misleading information*, masalah lain yang umum dijumpai dalam visualisasi data yang buruk ialah menyembunyikan informasi tertentu.
* **Distract**  
  Masalah lain yang mungkin terjadi ialah distraksi. Visualisasi data yang buruk sering kali menyertakan komponen visual yang seharusnya tidak dibutuhkan dan malah mengganggu proses pengambilan kesimpulan dari sebuah visualisasi data.

**Prinsip-Prinsip dalam Memvisualisasikan 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.

**Bercerita dengan Data**

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.