**Data analysis**

**Data analysis overview**

Analisis Data adalah proses pengumpulan, transformasi, pembersihan, dan pemodelan data.

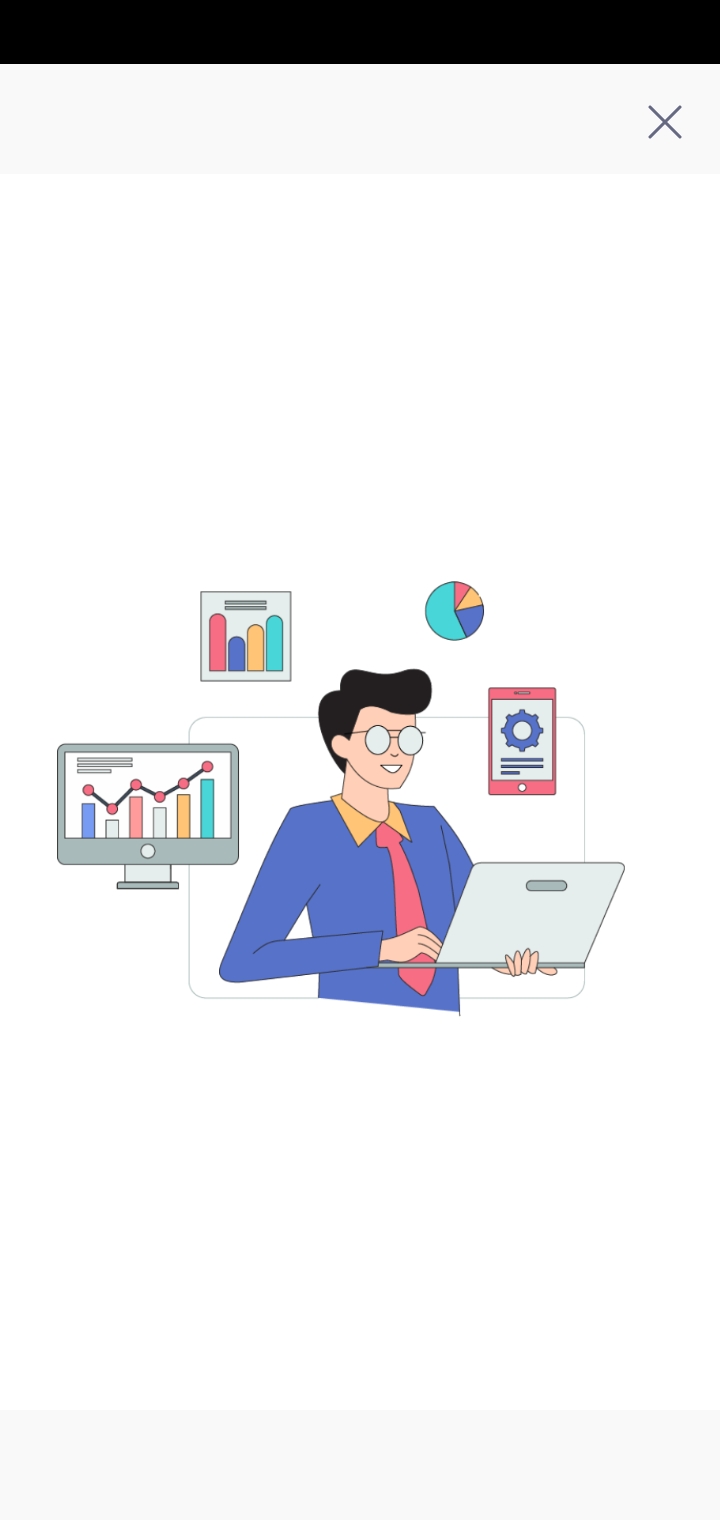
Tujuannya adalah menemukan informasi yang dibutuhkan.

Istilah Data modeling dan **Data analysis** memiliki arti yang sama.

Analisis mengacu pada pemecahan keseluruhan model menjadi komponen-komponen yang terpisah untuk diperiksa secara individual.

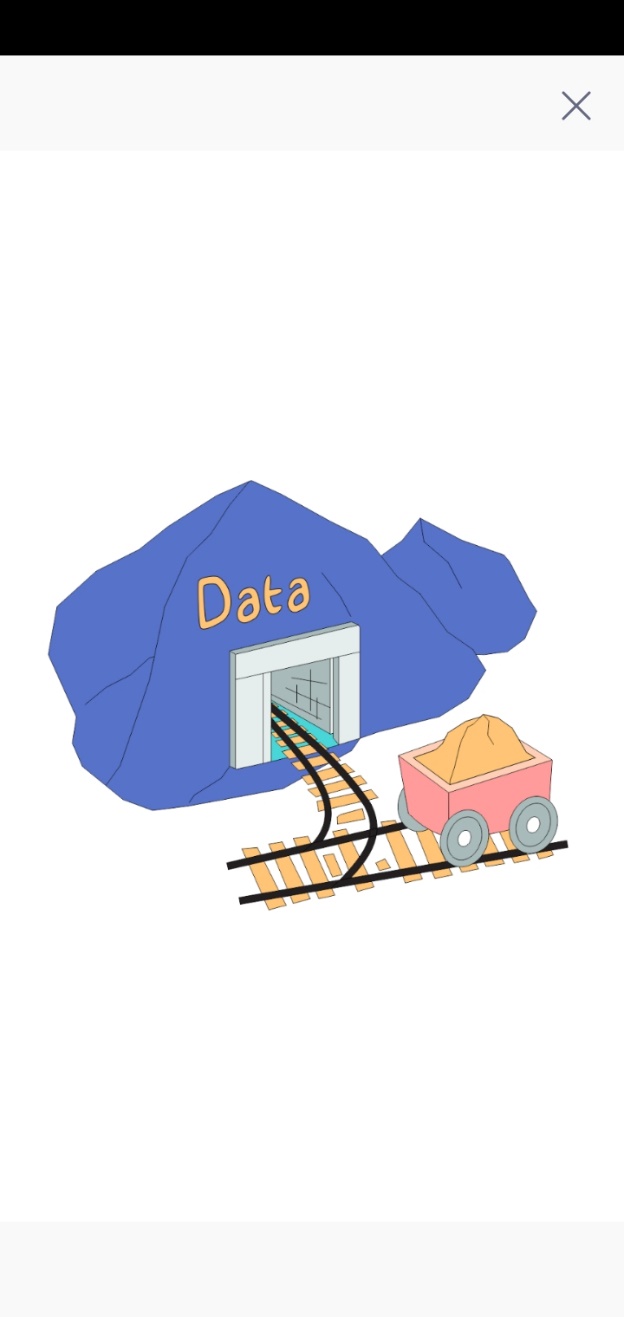
Analisis data adalah proses untuk mendapatkan data mentah dan mengubahnya menjadi informasi yang berguna untuk pengambilan keputusan oleh pengguna.

Data dikumpulkan dan dianalisis untuk menjawab pertanyaan, menguji hipotesis, atau membantah teori.

Ada beberapa teknik analisis data yang mencakup berbagai domain seperti bisnis, sains, ilmu sosial, dll.

pendekatan analisis data yang utama adalah:

* Data Mining
* Business Intelligence
* Statistical Analysis
* Predictive Analytics
* Text Analytics

**Data Mining**

adalah ***analisis data dalam jumlah besar***.

Hal ini dilakukan untuk mengekstrak pola data yang ***sebelumnya tidak diketahui dan menarik***.

Perlu diingat bahwa tujuannya adalah ekstraksi pola dan pengetahuan dari data dalam jumlah besar.

Analisisnya melibatkan metode ilmu komputer di persimpangan **artificial intelligence**, **machine learning**, **statistics**, **and database systems**.

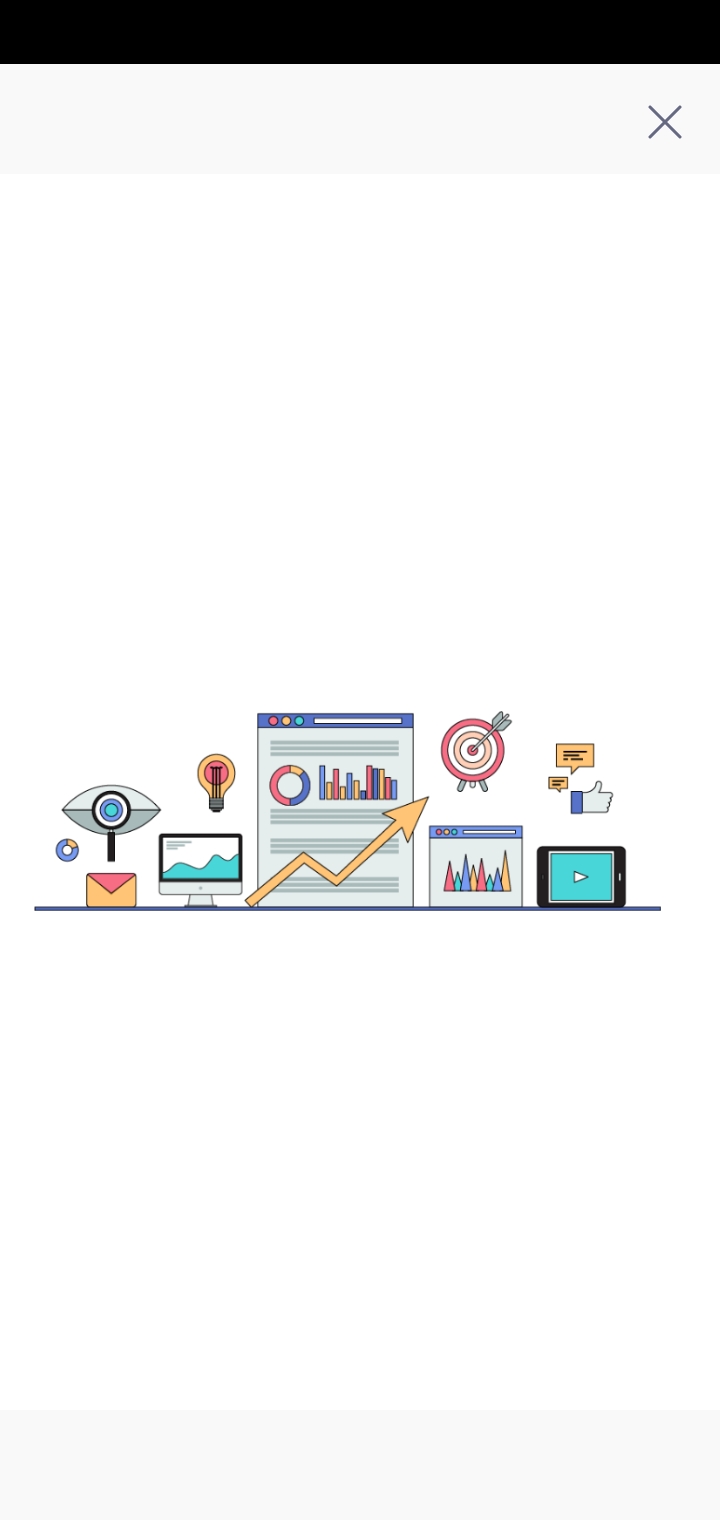
Pola yang diperoleh dapat dianggap sebagai ringkasan data masukan yang dapat digunakan dalam analisis lebih lanjut

**Business Intelligence**

Teknik dan alat bantu Business Intelligence digunakan **untuk akuisisi dan transformasi data tak terstruktur** dalam jumlah besar untuk membantu mengidentifikasi, mengembangkan, dan menciptakan peluang bisnis strategis baru.

Tujuan intelijen bisnis adalah untuk memudahkan interpretasi data dalam jumlah besar untuk mengidentifikasi peluang baru.

Hal ini membantu dalam menerapkan strategi yang efektif berdasarkan wawasan yang dapat memberikan keuntungan pasar yang kompetitif dan stabilitas jangka panjang bagi bisnis.



**Statistical Analysis**

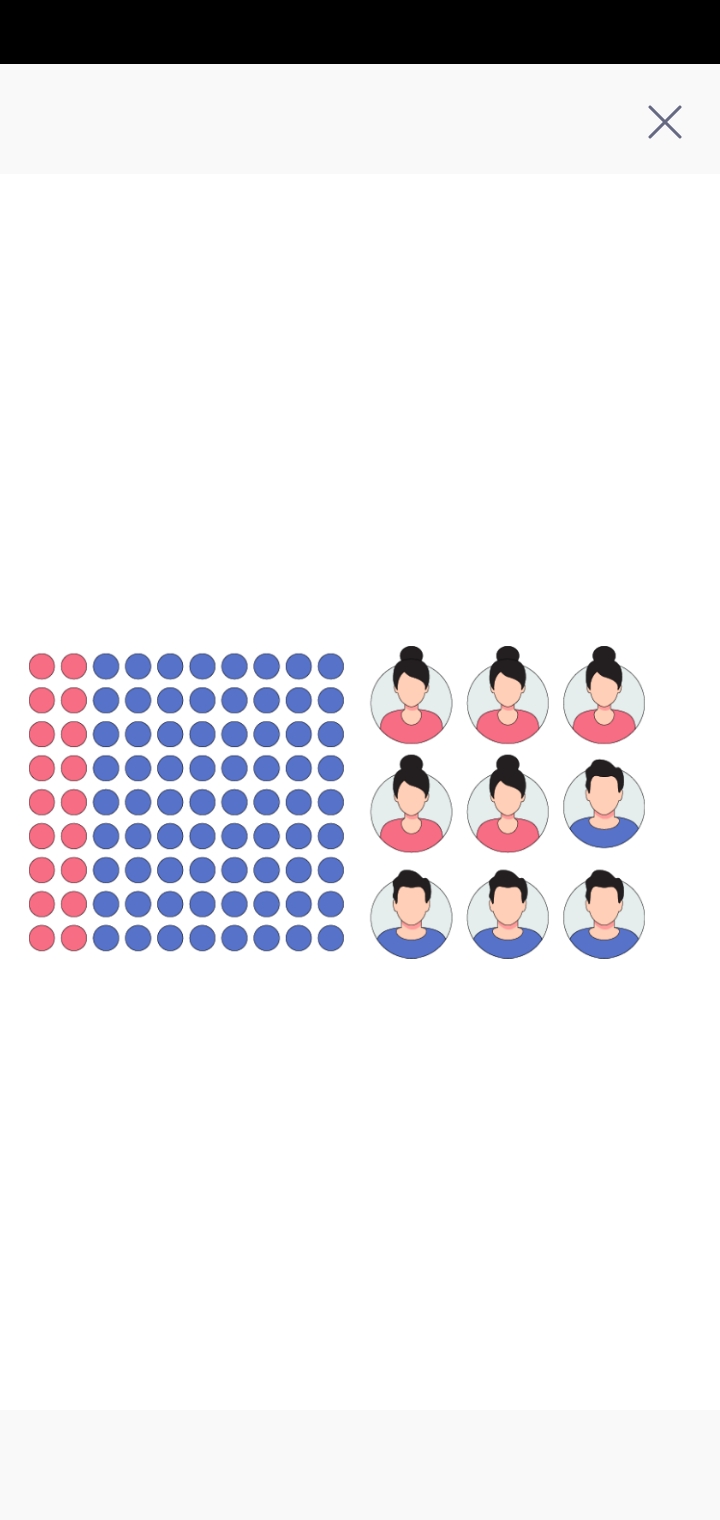
Statistik adalah studi tentang **pengumpulan, analisis, interpretasi, penyajian, dan pengorganisasian** data.

dalam analisis data, dua metodologi statistik utama digunakan –

**Descriptive statistics** - data dari seluruh populasi atau sampel dirangkum dengan deskriptor numerik seperti

* Mean, Standard Deviation for Continuous Data
* Frequency, Percentage for Categorical Data

**Inferential statistics** - statistik ini menggunakan pola dalam data sampel untuk memperhitungkan keacakan.

kesimpulan-kesimpulan ini:

* menjawab pertanyaan **ya**/**tidak** tentang data (pengujian hipotesis)
* memperkirakan karakteristik numerik dari data (estimasi)
* menggambarkan hubungan dalam data (korelasi)
* memodelkan hubungan dalam data (misalnya analisis regresi)

**Predictive Analytics**

**Flow of data anaysis**

**Let’s revise**