

Procesamiento de Datos

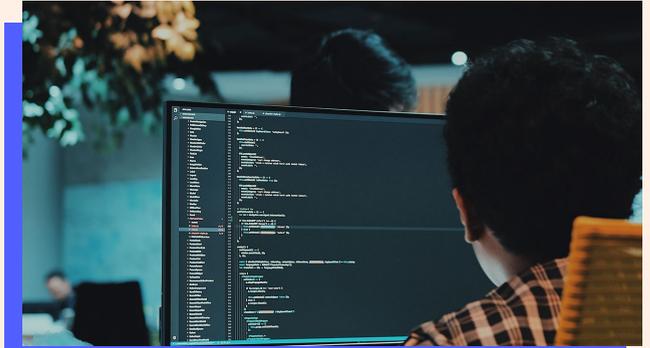
Cultura Digital II

Empezar

¿Qué es el procesamiento de datos?

El procesamiento de datos implica transformar datos sin procesar en información valiosa para las empresas. En general, los científicos de datos procesan datos, lo que incluye recopilarlos, organizarlos, limpiarlos, verificarlos, analizarlos y convertirlos en formatos legibles, como gráficos o documentos. El procesamiento de datos se puede realizar utilizando tres métodos, es decir, manual, mecánico y electrónico.

El objetivo es aumentar el valor de la información y facilitar la toma de decisiones. Esto permite a las empresas mejorar sus operaciones y tomar decisiones estratégicas oportunas. Las soluciones automatizadas de procesamiento de datos, como la programación de software de computadora, juegan un papel importante en esto. Puede ayudar a convertir grandes cantidades de datos, incluidos big data, en conocimientos significativos para la gestión de calidad y la toma de decisiones.



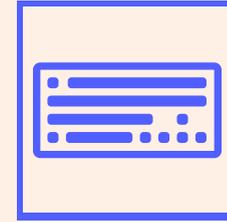
Etapas del ciclo de procesamiento de datos



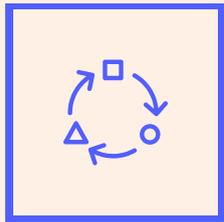
1. Recolectar datos



2. Preparación de datos



3. Datos de entrada



4. Procesamiento de
datos



5. Salida de datos



6. Almacenamiento de
datos



Esta etapa final del ciclo implica almacenar los datos procesados para uso futuro. Este paso se lleva a cabo después de usar la información requerida para implementaciones y conocimientos inmediatos. En esta etapa, las organizaciones almacenan datos con fines de referencia o para permitir un acceso fácil y rápido a los miembros de la organización para uso futuro.



En la etapa de salida de datos, también conocida como etapa de interpretación de datos, el procesador traduce y presenta los datos en un formato de datos legible, como documentos, gráficos, imágenes, etc. Ahora todos los miembros de la organización pueden utilizar los datos, y no solo los datos científicos, para ayudarlos en sus respectivos proyectos de análisis de datos.



Esta etapa implica el procesamiento de datos para su interpretación utilizando algoritmos de aprendizaje automático y algoritmos de inteligencia artificial. El proceso real puede diferir según la fuente de los datos (lagos de datos, redes sociales, dispositivos conectados) y su uso previsto o propósito (derivación de patrones y tendencias, determinación de soluciones o estrategias y optimización).



Esta es la etapa en la que los datos sin procesar comienzan a tomar forma informativa. Durante esta etapa, los datos limpios se ingresan en un sistema o destino (como un almacén de datos como Astera Creador de almacenes de datos o CRM como Salesforce). Esto se hace traduciéndolo a un idioma que el sistema pueda entender, ya sea manualmente o a través de dispositivos de entrada configurados para recopilar datos estructurados o no estructurados.

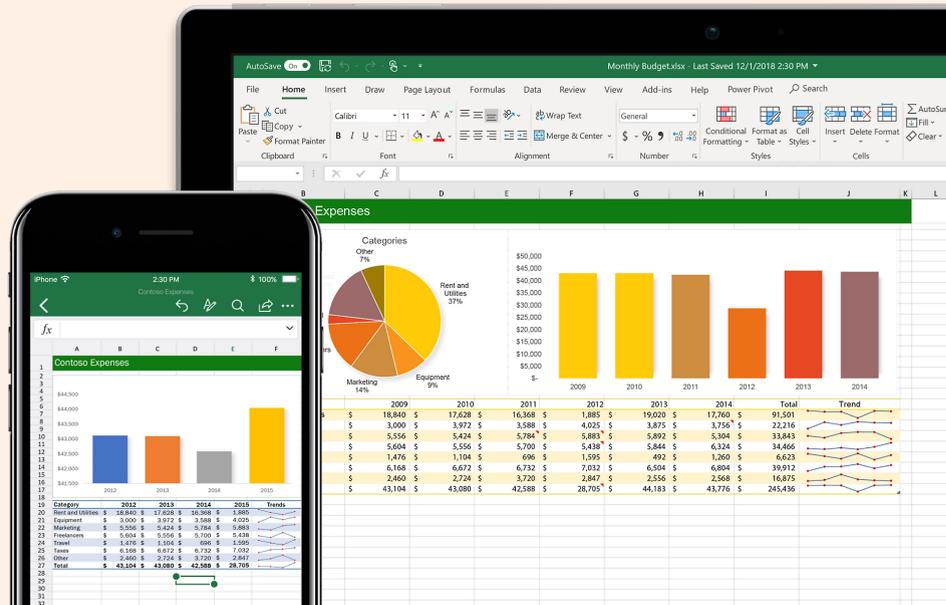


Los datos recopilados en la primera etapa luego se preparan y limpian. En esta etapa, también denominada "preprocesamiento", los datos sin procesar se organizan para ayudar en la implementación de etapas posteriores. La limpieza o preparación de datos implica la eliminación de errores, la eliminación de ruido y la eliminación de datos incorrectos (datos inexactos o incorrectos) para clasificarlos en datos de alta calidad.



Los datos se recopilan de fuentes confiables, incluidas bases de datos como lagos de datos y almacenes de datos. Es crucial que las fuentes de datos sean precisas, confiables y bien construidas para garantizar que los datos recopilados y la información recopilada sean de calidad y funcionalidad superiores.

Procesamiento de datos en la hoja de cálculo



Microsoft Excel es una fantástica herramienta para procesar datos, la estructura que adopta el programa en forma de hoja de cálculo permite analizar información a través de las celdas de Excel. Además, el programa pone a nuestra disposición diferentes herramientas para realizar un completo análisis en función del tipo de contenido que utilizamos.



Funciones estadísticas

Medidas de tendencia central

Las **medidas de tendencia central** son medidas estadísticas que pretenden resumir en un solo valor a un conjunto de valores. La principal utilidad de las medidas de tendencia central se relaciona a su capacidad de **sintetizar, resumir y representar**, con la mayor exactitud, un conjunto de datos en un solo número. Las medidas de tendencia central más utilizadas son: **promedio, mediana y moda.**

Función	PROMEDIO	MEDIANA	MODA
Descripción	Es la media aritmética y se calcula sumando un grupo de números y dividiendo a continuación por el recuento de dichos números.	Es el número intermedio de un grupo de números; es decir, la mitad de los números son superiores a la mediana y la mitad de los números tienen valores menores que la mediana.	Es el número que aparece más frecuentemente en un grupo de números.
Ejemplo fórmula	=PROMEDIO(A1:A20)	=MEDIANA(A2:A7)	=MODA(A1:A6)

Medidas de dispersión

Las **medidas de dispersión** consisten en números que otorgan información acerca de la variabilidad de los datos. Es decir, se encargan de mostrar **qué tan juntos o separados se encuentran los datos de una distribución**. Los valores de las medidas de dispersión nos permiten saber si los datos se encuentran estrechamente agrupados, si se encuentran ampliamente dispersos o si son iguales. Las medidas de dispersión son: **varianza, desviación estándar y coeficiente de variación**.

Función	VARIANZA	DESVIACIÓN ESTÁNDAR	COEFICIENTE DE VARIACIÓN
Descripción	Esta medida representa la variación que puede sufrir un conjunto de datos respecto a la media.	Corresponde a una desviación que es “habitual” entre el valor y la media. Se trata de un evento más probable y por lo tanto se emplea como tal en el cálculo de dispersión.	Se expresa como un porcentaje que es el resultado de dividir la desviación típica sobre la media del conjunto.
Ejemplo fórmula	=VAR(A2:A11)	=DESVESTA(A3:A12)	=DESVESTA(A1:A13)/PROMEDIO(A1:A13)



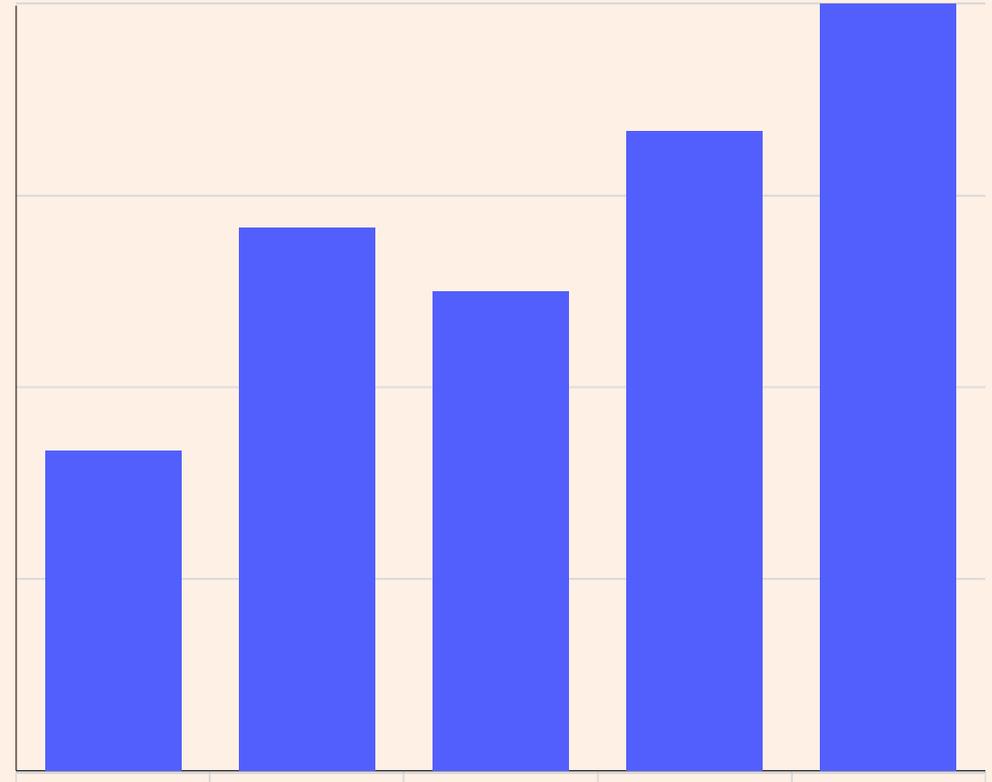
Gráficas

Gráficas

Un gráfico es una **representación gráfica de datos**.

La visualización de los datos por medio de gráficos ayuda a detectar **patrones, tendencias, relaciones y estructuras** de los datos.

Son de gran utilidad cuando se tienen que publicar estadísticas, comparar cantidades y expresar tendencias, dado que los datos que presentan son usualmente numéricos.



Tipos de gráficos en Excel

Columnas	Líneas	Circulares	Barras	Área	Dispersión
Un gráfico de columnas muestra normalmente categorías a lo largo del eje horizontal (categoría) y valores a lo largo del eje (valor) vertical.	Los gráficos lineales pueden mostrar datos continuos con el tiempo en un eje de escala regular y por tanto son idóneos para mostrar tendencias en datos a intervalos iguales.	Los gráficos circulares muestran el tamaño de los elementos de una serie de datos, en proporción a la suma de los elementos.	Este tipo de gráfico muestra comparaciones entre elementos individuales.	Los gráficos de área se pueden usar para trazar el cambio con el tiempo y para llamar la atención en el valor total en una tendencia.	Los gráficos de dispersión se utilizan por lo general para mostrar y comparar valores numéricos, como datos científicos, estadísticos y de ingeniería.

¡Gracias!