

# Lección 1: Conceptos básicos de ciencia de datos.



# Desarrollo de la sesión (4 horas)

## Análisis de datos



Los datos son números asignados de manera objetiva a una propiedad o característica de un objeto o un evento. Cuando nos referimos a una forma objetiva, nos referimos a que cualquier observador con un instrumento adecuado puede asignar ese mismo número a una misma propiedad o a un mismo evento. Por ejemplo, queremos determinar la altura de un árbol. Para esto podemos hacerlo con diferentes sistemas de medición, por ejemplo, podemos utilizar el sistema métrico internacional, con el cual asignamos un número a la altura del árbol, (por ejemplo 5 metros, que está en el sistema métrico internacional). Sin embargo, también podemos medirlo en sistema imperial (la medición se realiza en pies o yardas) o incluso podemos medirlo de forma no estandarizada, por ejemplo, medir la altura con algún elemento de referencia (medir con nuestra propia mano 15 manos de altura) lo cual también es una forma de medir, imprecisa, pero se acercaría al número asignado a la propiedad altura del árbol.

Encontramos, entonces que, las mediciones y conteos o datos nos permiten conocer propiedades de objetos (altura, ancho, peso, masa, entre otros) y las características de eventos (duración, tiempo de inicio, fechas, trabajo, energía, velocidades, aceleraciones, entre otros).



Por lo tanto, para conocer lo que nos rodea hemos ideado sistemas de medición que nos permiten, de manera indistinta asignar los mismos números a las mismas propiedades. Es allí donde seguimos estándares internacionales, que, para Colombia es el sistema internacional de unidades. Este sistema define muchas unidades de medida para que, de forma objetiva, en cualquier lugar del mundo podamos asignar las mismas cantidades a propiedades de objetos o características de eventos. En la tabla 1 se aprecian algunas unidades del sistema internacional.



Magnitudes físicas y unidades fundamentales del sistema internacional (SI)		
Magnitud física	Unidad	
	Nombre	Símbolo
Longitud	Metro	m
Masa	Kilogramo	kg
Tiempo	Segundo	s
Temperatura	grado kelvin	K
Intensidad de corriente	Amperio	A
Cantidad de sustancia	Mol	mol
Intensidad luminosa	Candela	Cd

**Tabla 1: Unidades base del sistema internacional de Unidades.**

Generalmente los datos vienen representados por una unidad de medida para conocer las magnitudes de forma estandarizada.

 En la última década se ha agregado un quinto nivel alineado con el desarrollo tecnológico asociado al big data (grandes cantidades de datos) y el aprendizaje de máquina, denominado inteligencia de negocios. En este nivel, se analizan grandes cantidades de datos generados de múltiples procesos, con el fin de tomar decisiones de alto nivel. Uno de los ejemplos apropiados es las máquinas de coca cola freestyle. Estos dispositivos representan un gran ejemplo del uso de big data en sistemas automatizados.





Las máquinas freestyle son dispensadores de bebidas que permiten a los usuarios mezclar bebidas gaseosas a partir de 15 sabores diferentes. En cada una de las máquinas se recolectaba la información de las ventas realizadas y de las preferencias de cada usuario para hacer las mezclas de bebidas. La empresa coca cola instaló más de 15.000 de estas máquinas en Estados Unidos y analizó como cambiaban los hábitos de consumo, determinando que la gente bebía mas los fines de semana y lo hacía menos en días festivos. Con estas máquinas se pudo determinar los hábitos de consumo en cada ubicación de cada tipo de bebida, por ejemplo, la Sprite se consumía mas en ciertos estados que la Fanta, y en el centro de Nueva York los dispensadores tenían más uso que en todo el país. Con lo que la empresa pudo instalar más y aumentar sus ventas. También encontraron que, en Georgia, la gente bebía más en horas de la noche en los restaurantes de 24 horas. S.



Pero el dato más importante es la mezcla más popular en todo el país. Se descubrió que la gente solía servir en sus vasos una mezcla de 33% Coca-Cola 33 % bebida de cereza y 34% Coca-Cola de vainilla. Con este dato se lanzó el sabor coca cola Cherry vainilla el cual en su primer año de lanzamiento vendió más de 2 millones de dólares. Todo esto a partir solamente de los datos de las máquinas freestyle. El punto de este caso de estudio es resaltar la importancia de los datos para la toma de decisiones de alto nivel y la inteligencia de negocios.



## ○ Análisis y exploración de los datos.



Con los datos es posible realizar muchos tipos de análisis que revelan información sobre patrones de comportamiento, temporadas, y mucha más información. Actualmente cualquiera puede recopilar muchos datos y tener analítica de fenómenos y procesos. Por ejemplo, hoy en día muchas personas cuentan con teléfonos inteligentes que vienen equipados con sensores de movimiento y de posición. De esta forma es fácil que cada dispositivo pueda dar seguimiento a la posición de su usuario, y analizando el movimiento con el tiempo es posible saber si una persona está caminando, va conduciendo un vehículo o va en un bote. Fusionando los datos de los sensores de posición con sensores de movimiento como acelerómetros y velocímetros, a través del análisis de las vibraciones detectadas en los sensores, es posible determinar si una persona va caminando, corriendo, saltando, si está en un gimnasio haciendo deporte, si está conduciendo una bicicleta e incluso si es el conductor o pasajero de un vehículo. Lo interesante es que todo esto sucede con la captura y analítica de datos y de algunos sensores. X



Para citar otro ejemplo, las organizaciones gubernamentales tienen la capacidad de validar en tiempo real si un vehículo es robado o no, analizando la placa con cámaras ubicadas en las calles, carreteras y semáforos. Incluso es posible determinar quienes cuentan con sus documentos vigentes y hacer un seguimiento de vehículos. Quienes ignoran que estos análisis existen suelen pensar que justo el día que tenían un documento vencido los detuvieron, pero en realidad, se debe a procesos de análisis de datos en tiempo real.



⊗

◉

◁ ▷

Teniendo en cuenta las posibilidades con los datos también es importante plantearnos ¿cómo podemos proteger nuestros datos? Ya que, si caen en manos equivocadas nuestra identidad, hábitos y muchos detalles personales pueden quedar al descubierto. Los datos son un activo valioso, y su manipulación responsable es fundamental. La capacidad de realizar análisis detallados que revelan patrones de comportamiento y proporcionan información sobre la posición y actividades de las personas es increíblemente poderosa. Sin embargo, esta misma capacidad, en manos equivocadas, puede convertirse en un riesgo para la privacidad y la seguridad individual.

⊗

◉

Suponga un mundo donde la información más íntima y personal cae en manos equivocadas. Desde detalles médicos hasta hábitos diarios, la exposición de estos datos podría tener consecuencias devastadoras. La seguridad de los datos no es simplemente un asunto técnico; es un componente esencial para salvaguardar la privacidad y la integridad de las personas. Además, las empresas y organizaciones que manejan grandes cantidades de datos están bajo una responsabilidad ética y legal para garantizar que la información confidencial de sus clientes y usuarios esté protegida contra amenazas cibernéticas.

Por este motivo, el análisis de datos no solamente se compone de algoritmos y programas que permiten la recolección y procesamiento masivo, sino que también hay que tener en cuenta qué implicaciones tienen los datos y qué hacen las empresas que recopilan nuestros datos.

Como actividad, cree una discusión con los estudiantes acerca de qué empresas recopilan nuestros datos y qué hacen con ellos.



**Analicen qué aplicaciones tienen acceso a la ubicación, fotos, llamadas, mensajes y otros datos que recopilan nuestros teléfonos. Ayude a los estudiantes a identificar qué podrían hacer con tales datos las compañías. Discuta con los estudiantes qué sucedería si la información que recopilan sus aplicaciones se hace pública. También oriente a los estudiantes a pensar como analistas de datos de grandes empresas y responda a las siguientes preguntas.**



¿Qué medidas de seguridad se deberían establecer para evitar las filtraciones de datos?

¿Qué límites deberían existir en la recolección de los datos?

¿Leen los usuarios de redes sociales y de aplicaciones móviles y web gratuitas los términos y condiciones cuando instalan?

¿Saben los usuarios qué hacen con sus datos?



Cierre la discusión concluyendo sobre las ventajas del análisis de datos, las posibilidades alrededor y las responsabilidades que cada persona tiene en el manejo de datos.