

Módulo 6

Sesgo, credibilidad, privacidad, ética y acceso. Organizar y proteger tus datos.

AGENDA DE LA CLASE

- ✓ Análisis de sesgo en datos
- ✓ Credibilidad de datos
- ✓ Ética de los datos
- ✓ Anonimización de datos
- ✓ Organización de los datos
- ✓ Protección de datos



01 - Análisis de sesgo en datos

El **sesgo de datos** son errores que sucede **sistemáticamente y que direcciona los resultados** transformando la información en datos poco confiables.

1

El **sesgo de la muestra** ocurre cuando una muestra **no es representativa** de la población en su conjunto.

2

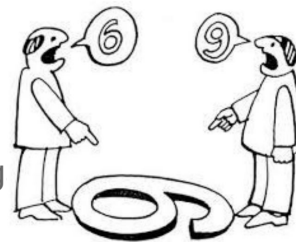
El **sesgo del Observador** es la tendencia de distintas personas a **observar las cosas de forma diferente**.

3

El **sesgo de interpretación** es la tendencia a **interpretar** siempre las **situaciones ambiguas** de manera positiva o negativa.

4

El **sesgo de confirmación** es la tendencia a buscar o interpretar la información de una manera que **confirma las creencias preexistentes**.



4
TIPOS

02 - Credibilidad de datos

Cuanto mayor sea la **calidad de los datos** que tenemos, mayor **confianza** tendremos en nuestras decisiones.

Los conjuntos de datos deben ser:



- **CONFIABLES**: Precisos, completos y no sesgados
- **ORIGINALES**: Validados con la fuente original
- **ABARCATIVOS**: Que contengan todos los datos necesarios para responder la pregunta o problema a solucionar
- **ACTUALIZADOS**: Que sea información vigente
- **CITADOS**: Tener referencias para citar su credibilidad

03 - Ética de los datos

La **ética de datos** se refiere a normas justificadas sobre **el bien y el mal** que dictan cómo se **recopilan, comparten y usan los datos**.

1 PROPIEDAD

- Las personas son **dueñas** de los datos que proveen y controlan su uso y procesamiento
- Esto responde a la pregunta: **¿quién es el dueño de los datos?**

2 TRANSPARENCIA DE LAS TRANSACCIONES

- Las actividades y algoritmos deben ser **explicables** y **entendidas** por los que **proveen los datos**
- Esto surge en respuesta a las **preocupaciones sobre los sesgos** de los datos

03 - Ética de los datos

3 CONSENTIMIENTO

- Derecho de las personas de **conocer cómo se va a usar sus datos** antes de compartirlo
- Deben conocer las respuestas a preguntas como:
 - ¿Por qué se recopilan los datos?
 - ¿Cómo se usarán?
 - ¿Durante cuánto tiempo se almacenarán?

4 VIGENCIA o utilidad financiera

- Derecho de las personas de conocer las **transacciones financieras resultado del uso de sus datos**.

03 - Ética de los datos

5 PRIVACIDAD

- **Preservar** los datos y la actividad de la persona cada vez que una transacción ocurre
- Todo gira en torno al **acceso, uso y recopilación** de datos
- Para las PERSONAS incluye el **derecho legal de una persona sobre sus datos**
- Para las EMPRESAS significa implementar **medidas de privacidad** para **proteger** los datos de las persona

6 APERTURA

- **Acceso, utilización y gestión libre**
- A veces nos referimos a estos como **datos abiertos** teniendo en cuenta los siguientes estándares:
 - La disponibilidad y al acceso
 - La reutilización y la redistribución
 - La participación universal

04 - Anonimización de datos

La **información de identificación personal**, o **PII**, es la información que se puede usar por sí misma o con otros datos para **rastrear la identidad de una persona**.

La **anonimización de datos** es el **proceso de proteger los datos privados** o confidenciales de las personas eliminando esa clase de información.



La **desidentificación**, que es un **proceso** utilizado **para eliminar toda la información de identificación personal** de los datos

04 - Anonimización de datos

¿QUÉ?

- Números telefónicos
- Nombres
- Patentes de vehículos y licencias de conducir
- Números de seguro social
- Direcciones IP
- Registros médicos
- Direcciones de correo electrónico
- Fotografías
- Números de cuentas



¿CÓMO?

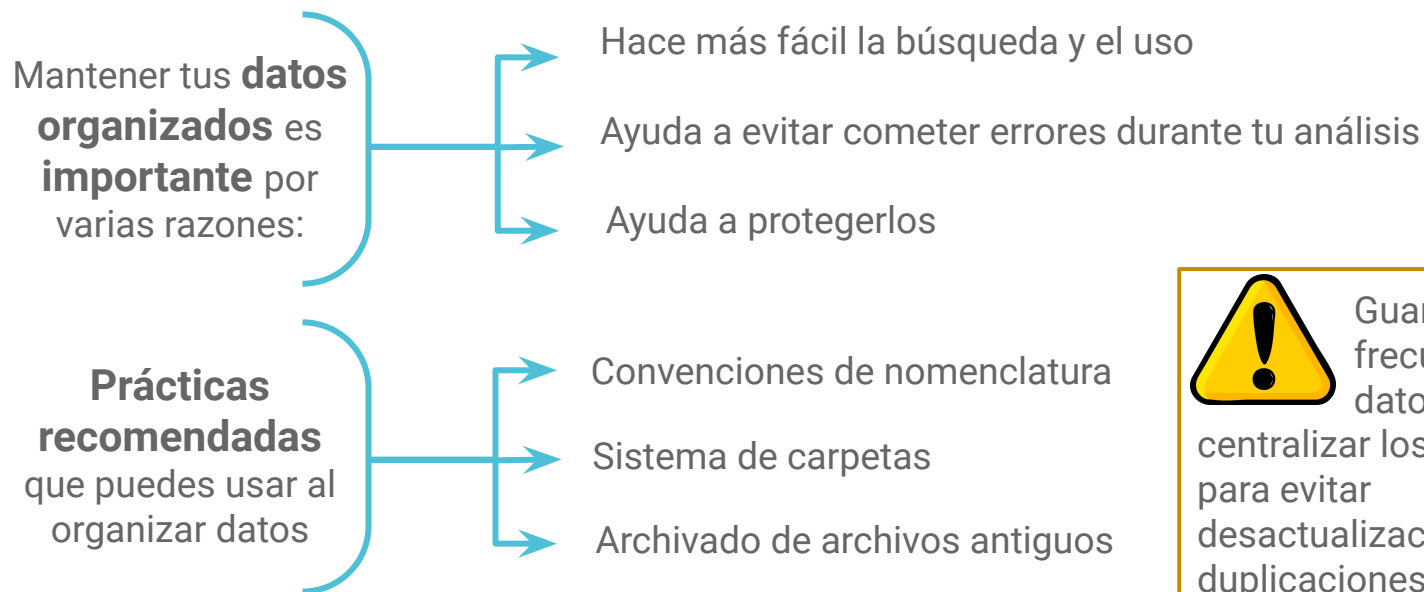
- Dejar en blanco
- Ejecutar el algoritmo hash
- Enmascarar la información personal

¿QUIÉN?

- Las **organizaciones** tienen la responsabilidad de proteger sus datos y la información personal que pudieran contener los datos.
- Como **analista de datos**, comprender qué datos deben anonimizar y exigir que se anonimicen los datos antes de trabajar con ellos.

05 - Organización de los datos

Gran parte del **proceso de preparación** de los datos para el análisis consiste en **organizar** y **proteger los datos**.



Guardar con frecuencia los datos y centralizar los mismos para evitar desactualizaciones o duplicaciones.

06 - Protección de datos

La **seguridad de datos** significa **proteger** los datos contra el **acceso no autorizado** o contra la **corrupción** adoptando medidas de seguridad

Por lo general, el propósito de la seguridad de datos es **impedir que usuarios no autorizados accedan a datos confidenciales** o los vean

Algunas medidas de seguridad que pueden ayudar a la protección de datos

El **cifrado** usa un algoritmo único para **alterar** los datos y hacer que los usuarios y las aplicaciones que no conocen el algoritmo puedan usarlos.

La **tokenización** **reemplaza** los elementos de datos que quieres proteger con datos generados aleatoriamente que se denomina "token". Los datos originales se almacenan en otra ubicación y se asignan en los tokens.



¿PREGUNTAS?





¡Muchas gracias!