

ANÁLISIS DE DATOS



Dato: unidad mínima de información.
Información: datos ordenados y organizados.
Conocimiento: conclusión, ideas, insights basados en el análisis.

DATA FOR GOOD

Datos para mejorar la calidad de vida de las personas.

- Responsabilidad en el análisis de los datos.
- Obtener muestras representativas y sin sesgos.
- Fairness.

EXPERIMENTOS EMPRESARIALES

Procedimientos que se realizan para apoyar, refutar o validar una hipótesis. Dan una idea sobre causa y efecto, con su demostración de resultados. Las empresas los pueden llevar a cabo con los clientes, y con sus datos, para mejorar el servicio. Se recolectan los datos de manera no convencional, y muchas veces puede resultar polémico.

INNOVACIÓN EN EL ANÁLISIS DE DATOS

Visión Estratégica:
Plan de acción a realizar.
Lineamientos de trabajo, especialmente aplicados en experimentaciones.
Visión Ética:
Responsabilidad, transparencia.
Visión Legal:
Actividades adecuadas a la normativa legal.

APLICACIONES DE WEB3

Metaverso:
Espacios virtuales 3D donde los usuarios interactúan en tiempo real en entornos creado con realidad virtual/aumentada

Bitcoin:
Protocolo de blockchain que revolucionó los pagos digitales. Algunos de sus protocolos sentaron las bases para las transacciones de criptomonedas.

NFT:
Non Fungible Token.



Tipos de Análisis



Describe patrones claves en los datos existentes y permite observar situaciones y comportamientos habituales.



Se definen conclusiones basadas en los datos históricos identificados en el análisis descriptivo.



Se identifican comportamientos futuros, con base a la información histórica.



Con simulaciones y optimizaciones se compara la información resultante del análisis predictivo, para implementar una acción.

Ventajas

- Mejor acceso para la toma de decisiones.
- Es posible generar gráficos y reportes ágilmente.
- Nuevo conocimiento generado de un conjunto de datos.

Desventajas

- Poca protección de datos (confidenciales y personales).
- Poca calidad de datos.
- Los datos no son todos iguales ni se generan de la misma forma.

HABILIDADES DE UNA ANALISTA DE DATOS

- Extraer, procesar y agrupar datos.
- Analizar esas agrupaciones de datos.
- Creación de informes y visualizaciones.
- Competencias matemáticas y estadísticas.
- Trabajo en equipo.
- Curiosidad por analizar e interpretar datos.
- Capacidad en la resolución de problemas.
- Responder incógnitas relevantes del negocio.
- Habilidades de comunicación.
- Habilidades analíticas y de negocio.

Asegurar la confianza en la tecnología

- Balancear la aplicación de IA al análisis de datos.
- Centrarse en la mejora de experiencia del cliente.
- Asegurar transparencia en el uso de los datos.
- Seguir buenas prácticas del uso de datos.

WEB1, WEB2, WEB3...

Web 1.0 (años 90)

- Red simple.
- Negocios en línea eran inexistentes.
- Portales de internet empiezan a cobrar importancia.
- Estructura centralizada.

Web 3.0 (presente)

- Busca solucionar los problemas presentados en Web 1.0 y Web 2.0, mediante la creación de una Internet descentralizada que sea de fácil acceso para todos, respetando privacidad y anonimato.
- Gracias a los avances de la Inteligencia Artificial, permite tener hoy máquinas que atienden en la web (chabots, etc).
- Se basa en blockchain (cadena de bloques). Red descentralizada construida sobre conexiones de igual a igual.

Web 2.0 (años 2000)

- La “Web Social”: participación colaborativa de los usuarios.
- Más personas se conectan y forman parte de la red.
- Contenido desarrollado por los usuarios.
- Estructura interconectada, reticular.

Blockchain

- Descentralizado.
- Transparente.
- Auditable.
- Redundancia de datos.
- Responsabilidad para los usuarios.

LOS DATOS Y LA WEB3

- Analítica de datos descriptica (ahora).
- Tareas de visualización, y análisis e investigación de Blockchain.
- Análisis computacional de la red.
- Los análisis y métricas son el alma de los ecosistemas blockchain.

VENTAJAS DE WEB3

- Mayor poder para los usuarios de internet.
- Descentralización de los servicios. Cualquiera puede ser propietario.
- Transparencia. Datos de Blockchain abiertos y públicos.

Recolección de los datos

DATASET

Colección de datos que puede verse en forma de “Tabla”.

REPRESENTACIÓN EN LA TABLA

Columnas (Features): características de los datos ejemplo: edad, sexo, fecha de nacimiento, profesión, temperatura.

Filas: cada muestra u observación dentro de la colección de datos.



ESTRUCTURA DE LOS DATOS

Datos estructurados: se representan en estructuras de tabla con filas y columnas. Puede contener números, fechas, categorías, cualidades etc.

Datos no estructurados: archivos de video, audio, imágenes emails.

Repositorios de datos:

[Machine Learning Repository](#)
[Kaggle](#)
[Imagenet](#)

TIPOS DE DATOS

Texto: Cadena de caracteres. Ejemplo: nombre, ciudad, cualidad.

Booleano: Representa la ocurrencia o no de un suceso. Puede codificarse con solo 0 y 1.

Numérico: representan normalmente cantidades o mediciones. Ejemplo: edad, estatura, precio, descuento.

NaN: No asignado/No se encuentra valor para este dato

TÉCNICAS DE RECOLECCIÓN DE DATOS

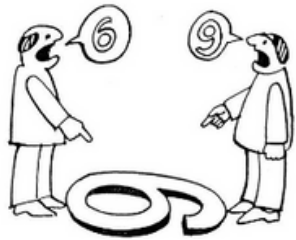
Métodos primarios: encuestas, entrevistas, focus groups, sondeos, métodos estadísticos.

Métodos secundarios: adquisición de bases de datos externas, web scraping, IoT, software CRM, informes de ventas.

Herramientas para recolectar: Formularios de google.

ESTRATEGIAS PARA LA RECOLECCIÓN DE DATOS

1. Definir el tipo de información que se desea recolectar y el objetivo.
2. Definir las herramientas de recolección de datos.
3. Sistematizar los datos recolectados (base de datos).



Sesgos (bias) en el dataset

SESGO (BIAS)

Es un defecto en el estudio de investigación científica en el método usado para recopilar o interpretar la información. Los sesgos pueden conducir a conclusiones incorrectas acerca de los resultados del análisis de datos.

El sesgo se puede aplicar en diferentes aspectos:

Estadística: La diferencia entre el valor esperado de un estimador y su estimado (estimado se entiende como el verdadero valor del parámetro a estimar).
Recopilación de datos: manipulación de la recopilación de datos para que su muestra no sea representativa de su población de interés. “Sesgo muestral”
Ciencias sociales y físicas: cualquier fenómeno que implica una influencia excesiva de condiciones pasadas -irrelevantes- a decisiones actuales.

TIPOS DE SESGO COMUNES

Sesgo de prejuicio: ocurre como resultado de influencias culturales o estereotipos.

Sesgo de la muestra: ocurre cuando los datos recopilados no representan con precisión el entorno en el que se espera que se ejecute el programa.

Sesgo de exclusión: ocurre como resultado de la exclusión de algunas características de nuestro conjunto de datos, generalmente bajo el criterio o la excusa de la limpieza de nuestros datos.

Sesgo del observador - sesgo del experimentador: se refiere a la tendencia a ver lo que esperamos ver, o lo que queremos ver.

Sesgo de medición: es básicamente la distorsión sistemática del valor ocurre cuando hay un problema con el dispositivo utilizado para observar o medir. Este tipo de sesgo tiende a sesgar los datos en una dirección particular.

TÉCNICAS PARA REDUCIR EL SESGO

¡Cuidado con los datos! no porque estén disponibles significa que estén listos para usar.

¡Equipos interdisciplinarios! buscar expertos en el tema

Dataset representativos con detalles relevantes

“Design Thinking” tener en mente al (humano) usuario final.

FEATURE ENGINEERING

(Ingeniería de las características)

Permite realizar un análisis más extensivo de lo que sería el aporte de las diversas características en el análisis de los datos.

EDA: Análisis Exploratorio de los Datos

EDA: Exploratory Data Analysis
Proceso crítico de realizar investigaciones iniciales sobre los datos para descubrir patrones, detectar anomalías, probar hipótesis y verificar suposiciones con la ayuda de estadísticas descriptiva y representaciones gráficas.



OBJETIVO DEL EDA

Hacer que los datos estén "limpios", ordenados, listos para entregar a un modelo, hacer un reporte. EDA se puede utilizar para eliminar o resolver todas las cualidades indeseables del dataset.

IMPORTANCIA DEL EDA

- Comprender la relación entre variables (columnas).
- Evaluar estadísticas del dataset.
- Identificar anomalías.
- Probar una hipótesis.
- Estimación de parámetros.

RETOS DEL EDA - ¿Cómo escoger qué variables explorar?

- Datasets con múltiples variables -> Identificación y reducción de variables.
- Conocimiento previo -> Preparación previa.
- Correlación de variables -> Análisis estadísticos de métricas.
- Manejo de tareas repetitivas -> Identificar diferencias, anomalías.
- Explorar datos no estructurados -> Herramientas adecuadas.
- Volumen de datos -> Herramientas adecuadas, reducción de dataset.
- Conclusión del EDA -> Tiempo, feedback y satisfacción de los interesados.

HERRAMIENTAS PARA EL ANÁLISIS EXPLORATORIO DE DATOS



PASOS DEL EDA

- **Recopilación de datos.** Recolección de información de manera sistemática.
- **Limpieza de datos.** Garantizar que los datos sean correctos y utilizables, al identificar errores o datos faltantes, corrigiéndolos y/o eliminándolos.
- **Preprocesado de datos.** Minería de datos. Transformar datos sin procesar en un formato comprensible: normalización, estandarización, selección de features.
- **Visualización.** Representación gráfica de información y datos. Gráficos estadísticos, diagramas, gráficos de información, storytelling.

TÉCNICAS DEL EDA

Univariante no gráfica. MEDIA-MEDIANA-MODA <ul style="list-style-type: none">- Una sola variable.- Sin relación.- Distribución de la población.	Univariante gráfica <ul style="list-style-type: none">- Una sola variable.- Boxplot, histograma, diagrama de tallo y hoja, gráfica de cuantiles normales.	Multivariante no gráfica <ul style="list-style-type: none">- Múltiples variables.- Relación entre dos o más variables.	Multivariante gráfica <ul style="list-style-type: none">- Múltiples variables.- Barchart, barplot, heatmap, bubble map, scatter plot o diagrama de dispersión.
--	---	--	--

Datos de Calidad

DATOS SUCIOS

- Outliers o valores atípicos.
- Valores faltantes (missing values).
- Valores nulos o NULL.

¿POR QUÉ SE GENERAN?

- Errores en entrada manual de datos.
- Información incompleta.
- Datos duplicados.
- Datos ambiguos.
- Errores en la transformación de datos.

CALIDAD DE DATOS

Factores para determinar si mis datos son de calidad, son confiables:

- Exactitud.
- Integridad.
- Actualización.
- Relevancia.
- Coherencia.
- Confiabilidad.
- Presentación apropiada.
- Accesibilidad.



GESTIÓN DE CALIDAD DE DATOS

- Volumen. Tamaño del dataset con el que se trabaja.
- Valor. De los datos generados.
- Velocidad. De procesamiento de los datos y la generación de reportes.
- Variedad. De tipos de datos.
- Veracidad. Integridad de los datos.

FALTA DE CALIDAD EN DATOS

- Genera:
- Falta de confianza del consumidor de los datos.
 - Pérdida de oportunidades de negocio.
 - Los arreglos demoran y resultan más caros.

TÉCNICAS PARA LOGRAR LA CALIDAD EN LOS DATOS

- Establecer Gobernanza de Datos.
- Identificar las funciones y las responsabilidades de los integrantes del Equipo de Datos.
- Crear las expectativas de calidad, y las estrategias empresariales de apoyo.
- Documentar las exigencias de Calidad de los Datos, definiendo las reglas para medir la misma,
- Implementar procesos de gestión de datos.
- Comprobar la validez de los datos en forma continua.



Tecnología para el análisis de datos

NOTEBOOKS ANALÍTICOS

Son cuadernillos o espacios donde podemos combinar código con nuestros datos.



Los más conocidos son Jupyter y Colaboratory, se pueden trabajar de manera local o en la nube.

¿Para qué sirven?

- Para procesar y estudiar enormes volúmenes de datos.
- Dejar notas y compartir proyectos con un equipo promoviendo el trabajo colaborativo.
- Desarrollar y analizar algoritmos.
- Sacar estadísticas y realizar gráficas, entre otras.

CONSIDERACIONES ÉTICAS

- Privacidad y seguridad de los datos
- Control de accesos
- Uso inapropiado de datos.
- Precisión en los datos(reales)
- Integridad de los datos.

TECNOLOGÍAS QUE SON HERRAMIENTAS EN EL ANÁLISIS DE DATOS

HOJA DE CÁLCULO

NUBE

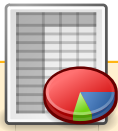
PROGRA-
MACIÓN

VISUA-
LIZACIÓN

BASE DE
DATOS

HOJA DE CÁLCULO

Espacio donde podemos ir definiendo datos en las diferentes filas y columnas, para ir manipulando, limpiando y sacando algunos datos estadísticos mediante gráficos. Las más populares son excel de microsoft y las spreadsheet de google esta última es una hoja de cálculo para trabajar en línea



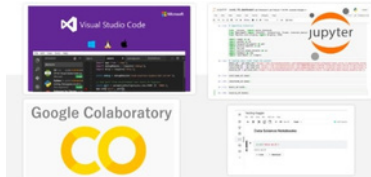
NUBE

La nube no es una entidad física, sino una red enorme de servidores remotos de todo el mundo que están conectados para funcionar como un único ecosistema (<https://azure.microsoft.com>)

Ventajas

- Escalabilidad
- Costos
- Disponibilidad

PROGRAMACIÓN



Dependerá de las necesidades del programador, del proyecto mismo, del lenguaje, y de la capacidad computacional,

VISUALIZACIÓN



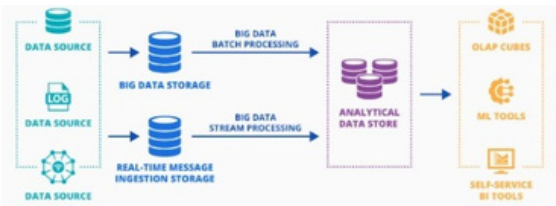
Dentro de las herramientas más conocidas podemos mencionar Tableau y Power BI.

BASES DE DATOS

Son estructuras más robustas donde almacenamos los datos.

TIPOS DE BASES DE DATOS:

- Las relacionales es donde se puede utilizar el lenguaje de consulta para poder extraer información, SQL(Structured Query Language), Access y Oracle. Me permiten definir tablas, hacer su esquema y establecer la relación entre ellas, encontrando información específica.



- NO relacionales, para el manejo de grandes volúmenes de datos.
 - Bases de datos relacionadas a columnas, a documentos y a grafos.
- Al manejar estas cantidades de bases de datos construimos el conocido Big Data.

R para el mundo de la estadística y Python para el mundo IA

LENGUAJE DE PROGRAMACIÓN

Es el mecanismo para comunicarnos con la computadora, compuesto de símbolos e instrucciones. Podemos clasificar los lenguajes de programación en 2:

1. De bajo nivel o lenguaje de máquina
2. De alto nivel o lenguaje humano

Los lenguajes de programación más conocidos para el análisis de datos son Python y R.

PYTHON

- Recolección de datos; CSV, JSON, Web
- Exploración de datos: Habilidades EDA (Pandas)
- Modelado de datos: Apoyo en bibliotecas como Numpy
- Visualizaciones: Menor fortaleza

VS

R

- Recolección de datos; CSV, JSON,
- Exploración de datos: Análisis estadístico y numérico más especializado
- Modelado de datos: Paquetes R, apoyo en Tidverse
- Visualización: Construido para análisis estadístico de datos



PYTHON

Lenguaje potente, fácil de aprender, estructura de datos a alto nivel es decir muy comprensibles y que aún así terminan siendo eficientes y simples. Su sintaxis y tipado es dinámico, su naturaleza es interpretada, lo cual es ideal para la elaboración de script y aplicaciones.

R

Bastante popular dentro de la academia, es decir dentro de investigaciones en universidades porque permite hacer la manipulación de los datos a un buen nivel para las personas que trabajan con estadística y matemática y por las herramientas de visualización.

Otros lenguajes:



5C, Consentimiento, Claridad, Consistencia, Confianza, Control y transparencia, Consecuencias

5 C

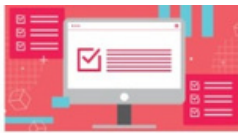
Consentimiento: Confianza y Convenio

Claridad: que los usuarios entiendan a que le están dando su consentimiento

Consistencia y confianza: Relación Persona-empresa donde el usuario ve la consecuencia de las políticas de protección de datos para depositar su confianza.

Control y Transparencia: Mayor control de los datos por parte del usuario para poder saber que hacen las empresas con nuestros datos.

Consecuencias: Dependiendo de las decisiones que hayamos tomado usando Data driven, tendremos consecuencias.



REGLAMENTACIÓN

GDPR= General Data Protection Regulation o RPDG (Reglamento General de Protección de datos)

Mencionando algunos derechos que establece esta nueva normativa destacamos:

- 1-Derecho a estar informado, básicamente proporcionar la transparencia sobre cómo van a ser utilizados estos datos personales.
- 2-Derecho al acceso, proveer acceso a los datos a cómo son suministrados y a cualquier información suplementaria que pueda ser utilizada junto a los datos.
- 3-Derecho a la rectificación, otorga el derecho que a los datos personales sean rectificados en caso de ser incorrectos o estar incompletos.
- 4-Derecho a ser borrado o también conocido como el derecho a ser olvidado, es el derecho a que los datos personales sean removidos.
- 5-Derecho a restringir el procesamiento, permite que los datos sean almacenados pero no procesados.
- 6-Derecho a la portabilidad de los datos, poder solicitar copias de la información almacenada sobre los datos personales.
- 7-Derecho a objetar, donde se le otorga el derecho a objetar sobre el procesamiento de los datos.
- 8-Derecho sobre la toma de decisiones, creación de perfiles automáticos, permite objetar sobre la toma de decisiones automáticas que se hagan con los datos.



Cultura Data-driven

CULTURA ÉTICA EN EL TRABAJO

Sistema de principios, valores y conductas que una organización adopta más allá del cumplimiento con la normativa legal obligatoria, para aprovechar los “conocimientos de los datos” y así identificar nuevas oportunidades de negocio, mejorar procesos, atender mejor a los clientes, entre otros.

¿EN QUÉ SE BASA UNA DECISIÓN DATA-DRIVEN?

Se basa en evidencia empírica que ayuda a los líderes a tomar sus decisiones, informados de la mejor forma.

COMITÉ DATA-DRIVEN

Comité multidisciplinario conformado por:

- Científicos o analistas de datos
- Científicos sociales
- Representantes de la comunidad.
- Expertos en privacidad de datos
- Defensores del consumidor



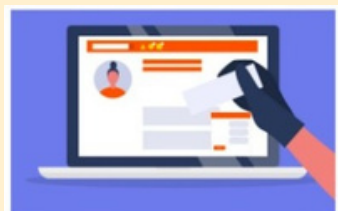
<https://www.google.com/search?q=comit%C3%A9+multidisciplinario+DATA&btn>

CARACTERÍSTICAS DE UNA EMPRESA DATA DRIVEN

- Datos: centralizados y ordenados.
- Política de Gobernanza de Datos : significa gestionar de la mejor manera posible la cantidad de información.
- Accesibilidad de los datos: la medida en que los datos se proporcionan en formatos abiertos y reutilizables.
- Herramientas de analítica: recursos tecnológicos que permiten a las empresas realizar análisis de una gran cantidad de datos en distintos contextos.
- Visualización datos: es la representación gráfica de información y datos. Proporcionan una manera accesible de ver y comprender tendencias, valores atípicos y patrones en los datos.

CONSIDERACIONES DEL DATA DRIVEN

- Está en línea con los valores de la organización.
- No viola ninguna normativa.
- Podría ser juzgado de forma negativa por un ciudadano promedio.
- Tiene beneficios que exceden los riesgos para la sociedad.
- No implica riesgos reputacionales o de polémica.
- Ethics-washing”, que significa que la empresa aparenta que tiene proyectos éticos poniendo a un comité que no funciona en la realidad o que tiene mucho miedo de discutir o contradecir.



Visualización de datos

¿QUÉ ES Y CÓMO NOS AYUDA?

Es la representación gráfica de la información y los datos. Nos ayuda a:

- Crear historias a través de datos
- Presentar los datos organizados y entenderlos con mayor facilidad.
- Exponer patrones, tendencias, correlaciones, etc
- Entender los datos y analizar un negocio con base a estos.
- Generar valor sobre lo construido con los datos y a la toma de decisiones.

BUENAS PRACTICAS DE VISUALIZACIÓN

- Elegir el gráfico adecuado y explicaciones intuitivas.
- Teoría del color.
- Tamaño y formas.
- Incluir descripciones: Títulos y leyendas.
- Uso de gráficos interactivos.
- Contar historia a partir de los datos.
- Evitar distracciones en gráficos y presentación (exceso de adornos, gráficos exagerados 3D)
- Evitar inconsistencia entre lo explicado y lo mostrado.

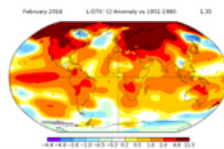
NIVELES DE VISUALIZACIÓN

Permite agrupar varias técnicas de visualización y se clasifican en 4 niveles:

Nivel 1: Visualización diseminativa

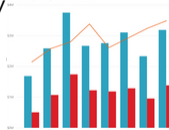
Los datos son usados para hacer representaciones sobre una ayuda.

Ej: mapa que muestra las temperaturas de una región.



Nivel 3: Visualización analítica

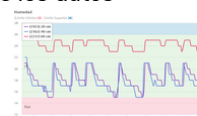
Permite examinar y entender relaciones entre diferentes variables.



Nivel 2: Visualización observacional

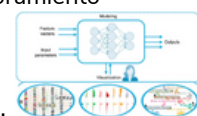
Permite la captura de los datos para la observación directamente.

Ej: mediciones de un sensor, datos en tiempo real.



Nivel 4: Visualización de desarrollo de modelos

Usado para el mejoramiento de modelos, algoritmos, métodos o sistemas existentes o para crear nuevos.



VENTAJAS

- Ayuda a entender o interpretar grandes volúmenes de datos.
- Apoyo en el aprendizaje como una ayuda visual.
- Más interesante, ayudando a la audiencia engancharse con el tema.
- Identificación más rápida de tendencias.
- Mejora la productividad.

DESVENTAJAS

- Falta de detalles.
- Puede ser propenso a malas interpretaciones
- Sin objetividad al transmitir el mensaje principal, por dar mucho enfoque a una introducción muy adornada.

Tipos de gráficos para representación de datos

¿CÓMO ELEGIR EL GRÁFICO ADECUADO PARA TUS DATOS?

La elección del gráfico más adecuado depende de qué quieras mostrar, qué preguntas quieres responder y a quién vas a presentar. Teniendo esto en cuenta, podemos clasificar algunos tipos de gráficos en 4 grandes casos de uso:

Comparación

Para comparar uno o más conjuntos de datos. Pueden comparar elementos o mostrar diferencias a lo largo del tiempo. Pueden comparar elementos o mostrar diferencias a lo largo del tiempo.



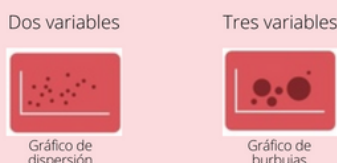
Composición

Para mostrar partes de un todo y cambios a lo largo del tiempo, cuando queremos saber cómo están compuestos estos datos o que características generales están presentes en los conjuntos de datos.



Relación

Para mostrar una conexión o correlación entre dos o más variables.



Distribución

Para mostrar cómo se distribuyen las variables a lo largo del tiempo, lo que ayuda a identificar valores atípicos y tendencias.



[Bibliografía](#)

PROFUNDICEMOS EN ALGUNOS TIPOS DE GRÁFICOS...

Gráfico de línea

Ayuda a mostrar los datos sobre el tiempo, entender tendencias e identificar intervalos. Puede contener 1 o más variables.



Gráfico de barras/columnas

Ayuda visualizar comparaciones numéricas de diferentes categorías. Los datos están asociados a los conceptos que se valoran y por lo tanto se exponen a información cuantitativa.



Histograma

Permite hacer una distribución de datos a través de un período o intervalos de tiempo. Permiten estimar la concentración de los datos, valores extremos e inusuales.

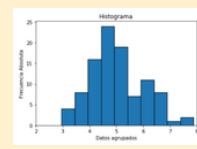


Gráfico de dispersión

Identifica si existe una relación o correlación entre dos variables. Tipos de correlación:

- **Positiva:** los valores aumentan juntos (imagen)
- **Negativa:** un valor disminuye mientras el otro aumenta
- **Sin Correlación.**

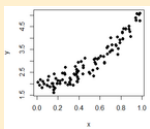


Gráfico de cajas y bigotes

Es la representación estadística de 5 valores (los tres cuartiles y los mínimo y máximo de los datos). Y ayuda a entender la variabilidad de los datos.

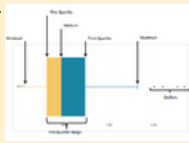


Gráfico de área

Usados para mostrar el desarrollo de valores cuantitativos durante un intervalo de tiempo, mostrar ciertas tendencias y además podrían estar apilados uno encima de otros de ser necesario.



Mapas de calor

Permiten visualizar datos a través de variaciones de color. Útiles para mostrar diferencias mediante múltiples variables. Algo importante al usarlo son las leyendas, es decir indicar qué representa cada color.

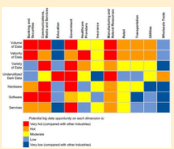


Gráfico de barra apilada

Permite comparar valores numéricos con respecto a ciertas categorías. Es importante tener en cuenta la diferencia de colores y sus respectivas leyendas.

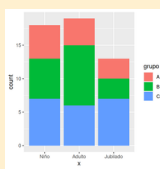
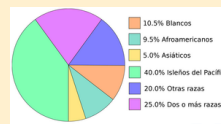


Gráfico de torta/circular

Muestra de un total la división de categorías. Es importante la representación de los colores y las etiquetas y recordar que se trabaja bien cuando se manejan pocas categorías.



Data Storytelling

¿QUÉ ES?

Es una técnica que utiliza datos para contar una historia. Permitiendo organizar todos los datos y presentarlos de una manera agradable y accesible. Es importante pasar por las siguientes fases:



HABILIDADES QUE COMPONEN EL DATA STORYTELLING

El data storytelling se fundamenta en:

Datos + Visualización + Narrativa

A estos pilares se le puede sumar el **Contexto de comunicación**, que se basa en habilidades que permiten un buen desarrollo de la historia a contar y llegar mejor a la audiencia. Cada intersección entre los pilares permite llegar a un resultado distinto y de mucho valor para la entrega de los datos.



IMPORTANCIA

- Ayuda a dar significado.
- Diferenciador en la competencia.
- Impulsa la comunicación y el compromiso de los clientes.
- Habilidad para profesionales de datos.
- Impacto en métricas.
- Que sea escalable.

CÓMO TRABAJARLO...

- ¿Qué quiero informar?
- ¿Cuáles son los resultados importantes?
- ¿Cuál es la historia que me permitirá contar los resultados?
- ¿Quiénes me escucharán?
- ¿Cuál será mi llamado a la acción?

BUENAS PRACTICAS

- Narrativa: inicio - nudo - desenlace.
- Iniciar con una pregunta.
- Terminar con un llamado a la acción.
- Apoyo en la presentación con gráficos.
- Hablar sin mucha complejidad.
- Tener en cuenta el objetivo de la presentación.
- Conocer tu audiencia.
- Control de tiempo.

Privacidad y visualización de datos

USO DE LOS DATOS

Después del proceso de análisis de datos, es importante preguntarnos: Al hacer uso de los datos, ¿cómo una empresa debe administrar sus datos para resguardar la privacidad de sus datos sensibles?



PRIVACIDAD Y SEGURIDAD DE DATOS

Aspectos importantes para la gobernanza de datos



VISUALIZACIÓN DE DATOS

Interacción humano-data: Interfaz interactiva entre un ser humano y la representación visual de los datos, análoga a la interacción humano-computadora.

Todo lo que veamos con relación al manejo de los datos, tiene una influencia directa a los aspectos relacionados en la imagen.



GOBERNANZA DE DATOS

¿Qué es?

Es el ejercicio de tomar decisiones y autoridad para asuntos relacionados con los datos.

La gestión de datos necesita una estrategia y ser sostenible en el tiempo en la organización.

Ventajas

- Confianza y mitigación de riesgos.
- Difusión de conocimientos.
- Apoyo en los programas empresariales.
- Calidad de datos.
- Gestión de metadatos.
- Almacenamiento de datos.
- Seguridad de los datos.

Principios

- Integridad
- Transparencia
- Auditabilidad
- Responsabilidad
- Gestión
- Control
- Estandarización
- Gestión de Cambio

Buenas prácticas

- Identificación temprana de datos críticos.
- Tratar los datos como recursos estratégicos.
- Establecer políticas y procedimientos para el ciclo de vida de los datos.
- Involucrar usuarios en el proceso.

ANONIMIZACIÓN DE DATOS

¿Qué es?

Proceso para codificar/encryptar datos sensibles de un conjunto de datos. Fomentando así un movimiento de datos más seguro entre sistemas.

Tipos

- K-anonimización
- Anonimización determinista
- Anonimización probabilista