

INTRODUCCION A LA CIENCIA DE DATOS CON PYTHON

Unidad 1:

CIENCIA DE DATOS

Unidad 1: CIENCIA DE DATOS

OBJETIVOS

- Comprender el **significado de la ciencia de datos** y su importancia en diversos campos.
- Familiarizarse con la **terminología y los conceptos básicos** de la analítica de datos.
- Identificar los **roles comunes** dentro del proceso de ciencia de datos y sus responsabilidades.
- Entender las diferentes **fases del proceso** de ciencia de datos, desde la obtención de datos hasta la presentación de resultados
- Aprender sobre la importancia de la **obtención de datos de calidad y la ingeniería de características** para el éxito de un proyecto de ciencia de datos.
- Entender la importancia de la **disponibilidad de datos** para las organizaciones y proyectos de ciencia de datos.
- Reconocer la importancia de la **gobernanza de datos en la gestión** y el uso responsable de la información.
- Comprender los **desafíos éticos** asociados con el uso de datos en proyectos de ciencia de datos.

CONTENIDOS

- Ciencia de datos: definición y su importancia en la actualidad. Aplicaciones prácticas en el campo de las ciencias económicas.
- Introducción a la analítica de datos (*Data Analytics*): terminología y conceptos fundamentales, categorías de *Data Analytics*: descriptivo, predictivo, prescriptivo.
- El proceso de data science: roles comunes, datos, obtención de datos, curación e ingeniería de características (feature engineering).
- *Big data*: introducción y nociones básicas. Identificación de grandes volúmenes de datos
- Arquitectura de datos.
- *Business Intelligence*.
- Disponibilidad de datos
- Gobernanza de los datos
- Ética en la ciencia de datos.
- Software de programación en ciencia de datos

Ciencia de datos: definición y su importancia en la actualidad. Aplicaciones prácticas en el campo de las ciencias económicas

La **toma de decisiones basada en datos** está cambiando nuestra forma de trabajar y vivir.

Desde la ciencia de datos y el aprendizaje automático hasta la analítica avanzada y los cuadros de mando en tiempo real, los responsables de la toma de decisiones demandan datos precisos y oportunos que les permitan tomar decisiones informadas.

La **gestión de datos** adquiere una importancia crucial en un entorno donde la toma de decisiones depende cada vez más del acceso a información precisa, actualizada y estructurada.

La **Ciencia de Datos** es un campo interdisciplinario que combina estadística, matemáticas, informática y conocimiento del dominio para extraer información valiosa a partir de datos. Su objetivo es transformar grandes volúmenes de datos en conocimientos accionables mediante técnicas de análisis, visualización y modelado.

Introducción a la analítica de datos (Data Analytics): terminología y conceptos fundamentales, categorías de Data Analytics: descriptivo, predictivo, prescriptivo.

La **Analítica de Datos** es el proceso de examinar, limpiar, transformar e interpretar datos con el objetivo de obtener información útil, tomar decisiones informadas y predecir tendencias. Se utiliza en múltiples campos, como negocios, educación, salud y ciencias sociales.

Analíticas descriptivas: ¿Qué pasó?

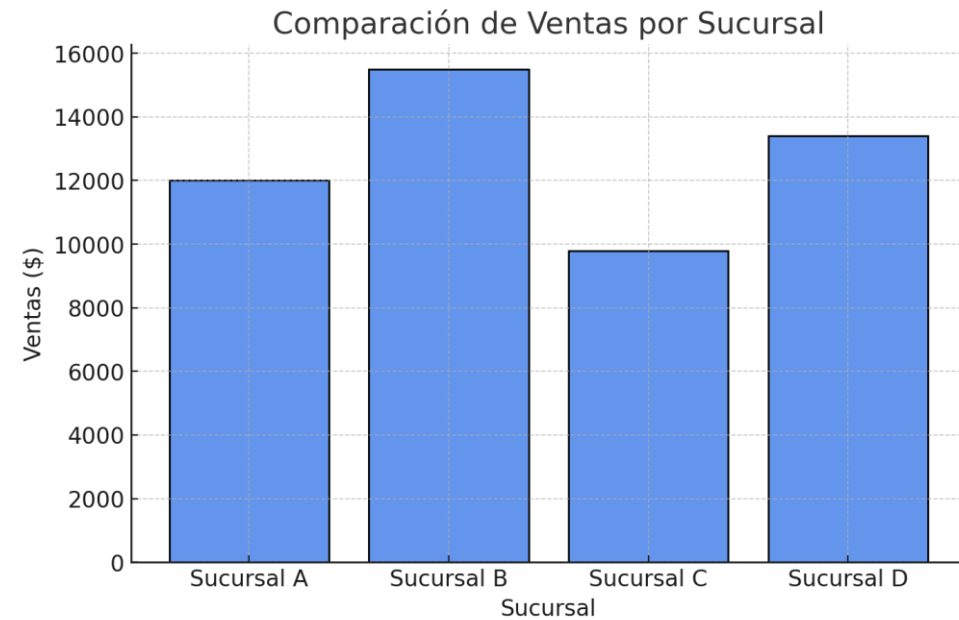
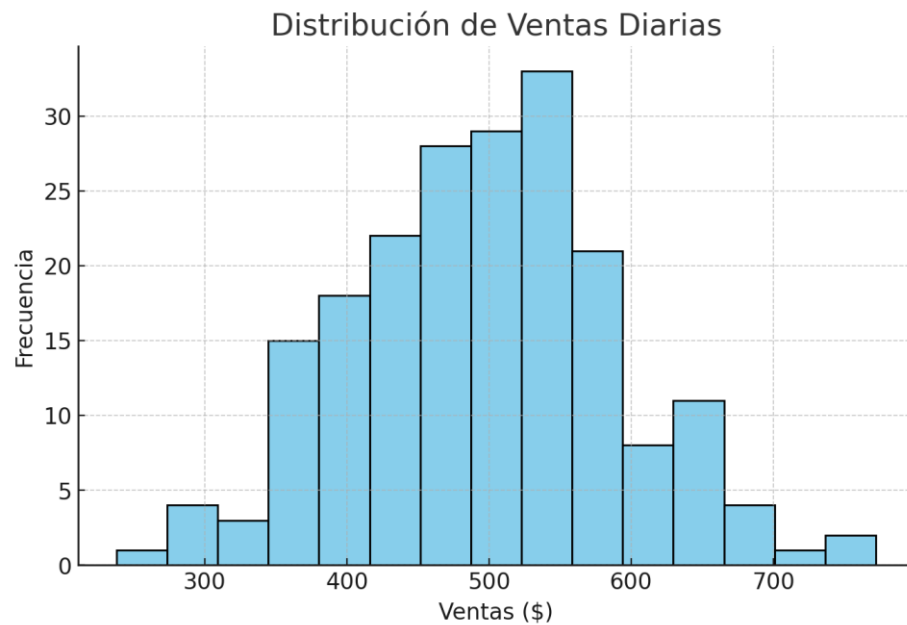
Analíticas predictivas: ¿Qué puede pasar?

Analíticas prescriptivas: ¿Qué debería pasar para influir en el futuro?

Data Analytics

Analíticas descriptivas

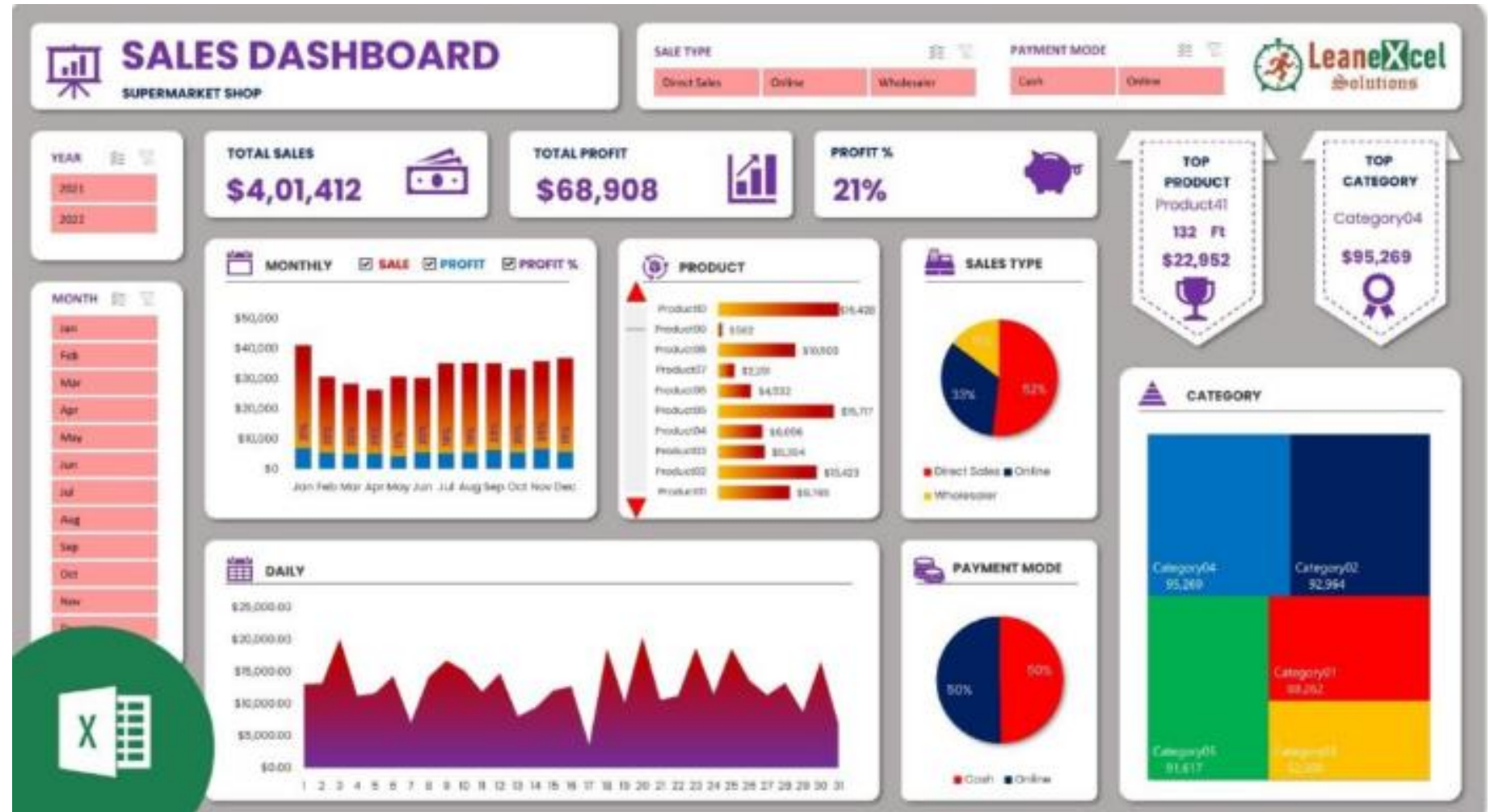
- Resumen datos para **identificar patrones** que podrían proporcionar nuevos conocimientos a los responsables de la toma de decisiones.
- Permiten que los responsables de la toma de decisiones "**desglosen**" los datos en detalle.



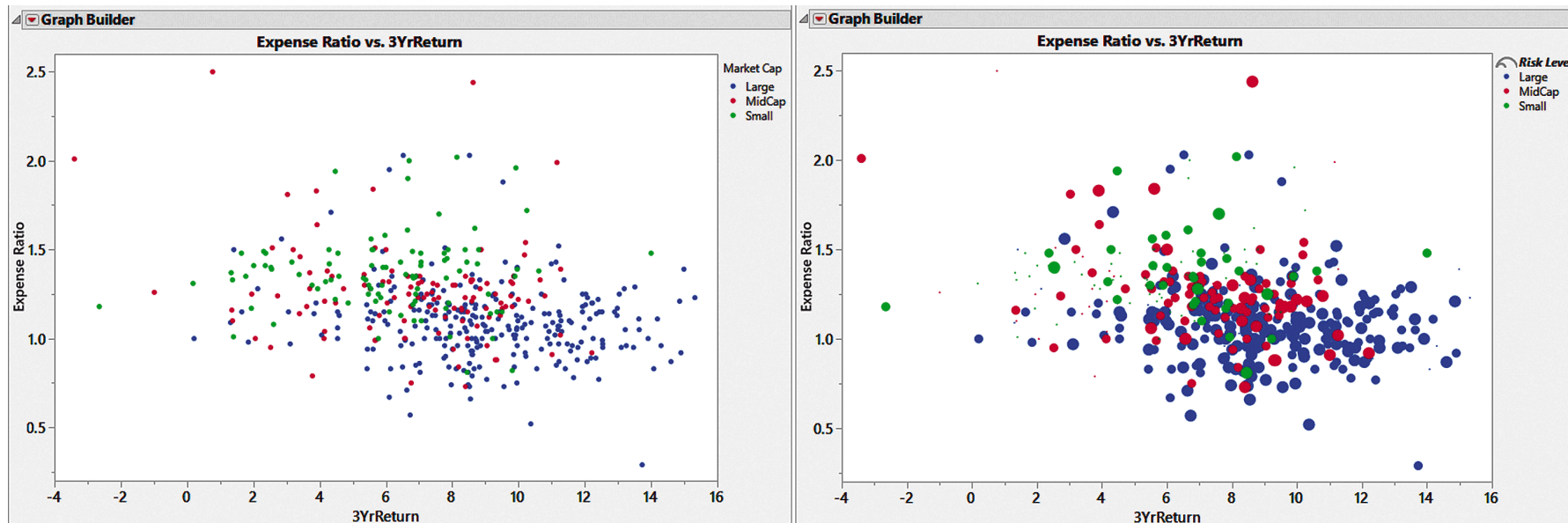
Data Analytics

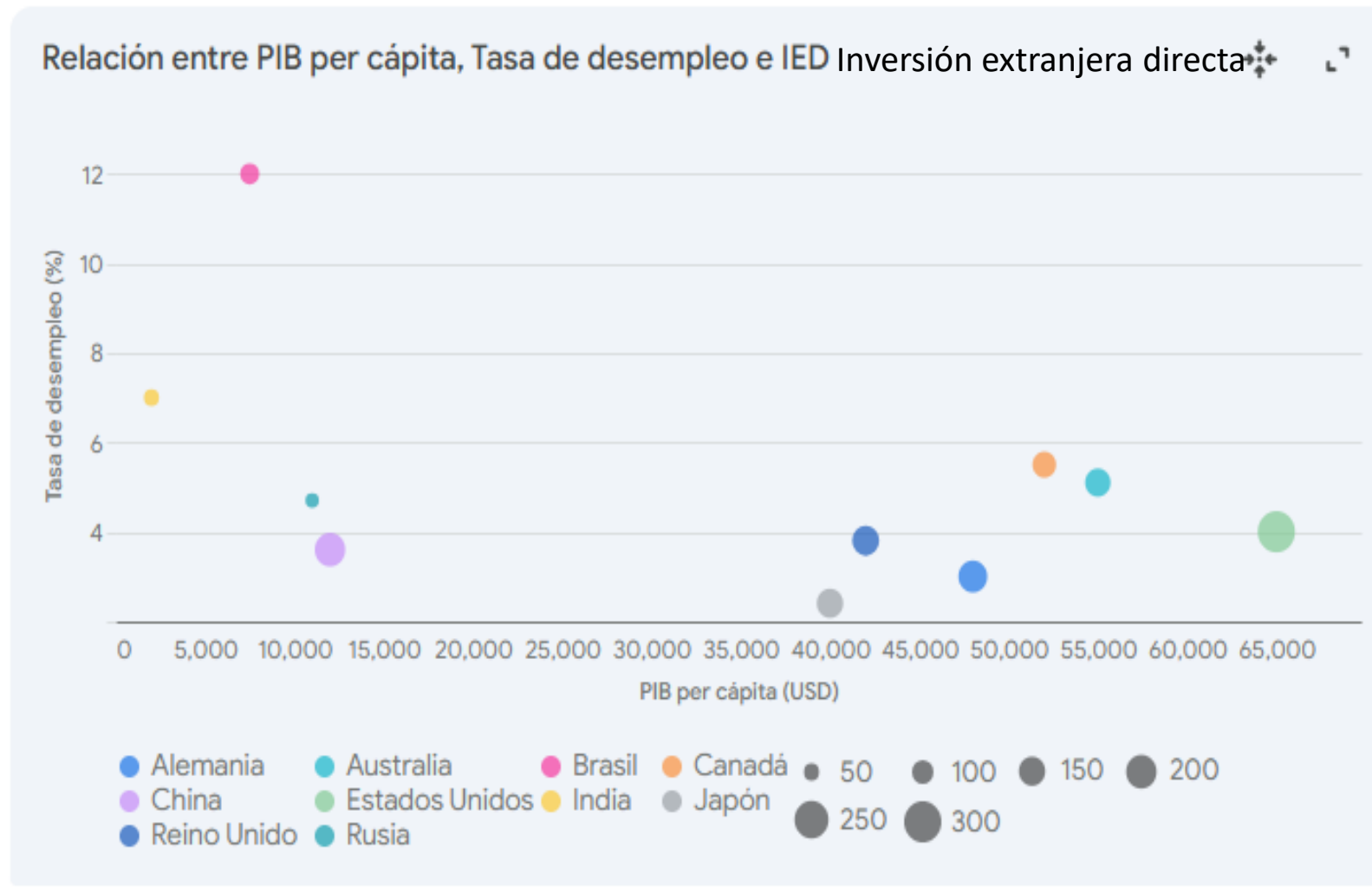
Analíticas descriptivas

Tableros (Dashboards)



- Las analíticas descriptivas a menudo tienen una **alta dimensionalidad** de los datos: capacidad de visualizar un gran número de variables.
- Las visualizaciones **utilizan el color, el tamaño o el movimiento** para representar varias dimensiones adicionales.
- Ejemplo: las **gráficas de dispersión** bidimensionales siguientes utilizan el **color** para mostrar una tercera dimensión.





Data Analytics

Analíticas predictivas






- Los **métodos de predicción** utilizan datos para predecir valores objetivos. Por ejemplo, la probabilidad de que se produzca un evento.
- Los **métodos de clasificación** asignan elementos de una muestra a categorías o clases de destino.
- Los **métodos de agrupación en clústeres** encuentran agrupaciones en los datos que se están analizando.
- Los **métodos de asociación** encuentran elementos que tienden a ocurrir juntos.

- **El análisis predictivo** a menudo se basa en métodos estadísticos inferenciales como la regresión.
- Los responsables de la toma de decisiones que utilizan el análisis predictivo deben comprender cómo se utilizan los **métodos estadísticos inferenciales**.
- El uso de ***Data Analytics*** sin un adecuado conocimiento de estadísticas inferenciales puede llevar a interpretaciones erróneas y decisiones perjudiciales para cualquier responsable de la toma de decisiones. Sin una base estadística sólida, los análisis pueden ser sesgados, irrelevantes o incluso engañosos, afectando negativamente la estrategia y el rendimiento organizacional.

Data Analytics

Analíticas predictivas: métodos supervisados vs. no supervisados

- Los **métodos supervisados** construyen modelos partiendo de datos etiquetados previamente.
 - **Datos etiquetados:** conocemos el valor de los atributos que son nuestro objetivo.
 - **Ejemplo:** un modelo para predecir la cesación de pagos de clientes.
- Los **métodos no supervisados** crean modelos sin datos de entrenamiento.
 - Trabajan **sin datos etiquetados** previamente.
 - Se parte de **estructuras no evidentes** subyacentes en los datos.
 - **Ejemplo:** segmentación de los clientes en grupos homogéneos.

Métodos		<u>Predicción</u>	<u>Clasificación</u>	Clustering	<u>Asociación</u>
Supervisados	Árboles de Regresión				
	Árboles de Clasificación				
No supervisados	Análisis de Clúster				
	<u>Escalado Multidimensional (MDS)</u>				
	Análisis de Correspondencias Múltiples (MCA)				

- Los análisis prescriptivos buscan:
 - Recomendar la mejor acción a seguir para lograr un resultado deseado.
 - Optimizar el rendimiento de una empresa.
 - Hacer recomendaciones sobre cómo responder y gestionar las circunstancias futuras del negocio.
- El análisis prescriptivo combina:
 - Métodos estadísticos tradicionales.
 - Herramientas de las ciencias de la administración.
 - Sistemas de información.
- Algunos métodos de análisis prescriptivo:
 - Optimización, simulación y análisis de texto.

El proceso de *data science*: roles comunes, datos, obtención de datos, curación e ingeniería de características (*feature engineering*).

Roles en proyectos de ciencia de datos

- *data scientist*
- *data architect*
- *data engineer*
- *data business analytics*
- *IT/infraestructura engineer*
- *Usuario*

El proceso de *data science*: roles comunes, datos, obtención de datos, curación e ingeniería de características (*feature engineering*).

Data Scientist

- **Visualización de datos:** Crear gráficos, *dashboards* y representaciones visuales para comunicar insights de manera clara y efectiva.
- **Machine Learning:** Desarrollar, entrenar y optimizar modelos predictivos y algoritmos de aprendizaje automático para resolver problemas.
- **Análisis avanzado:** Explorar datos para identificar patrones, tendencias y oportunidades.

Interacción con otros roles:

Trabaja con **Data Engineers** para acceder a datos limpios y preparados.

Colabora con **Data Business Analysts** para entender las necesidades del negocio.

Proporciona visualizaciones y modelos útiles para los **Usuarios**.

El proceso de *data science*: roles comunes, datos, obtención de datos, curación e ingeniería de características (*feature engineering*).

Data Architect

- **Diseño de sistemas de datos:** Crear la estructura y arquitectura de bases de datos, data warehouses y data lakes.
- **Mantenimiento:** Asegurar que los sistemas de datos sean escalables, seguros y eficientes.
- **Integración de datos:** Conectar diferentes fuentes de datos para garantizar la coherencia y accesibilidad.

Interacción con otros roles:

Proporciona la infraestructura de datos a **Data Engineers** y **Data Scientists**.

Asegura que los sistemas cumplan con los requisitos del negocio, en colaboración con **Data Business Analysts**.

El proceso de *data science*: roles comunes, datos, obtención de datos, curación e ingeniería de características (*feature engineering*).

Data Engineer

- **Disponibilidad de datos:** Garantizar que los datos estén accesibles, limpios y listos para su uso.
- **Ingeniería de datos:** Construir y mantener pipelines de datos para recopilar, transformar y almacenar información (conjunto de procesos automatizados que transportan datos desde una o varias fuentes hasta su destino final,).
- **Automatización:** Crear procesos automatizados para la recolección y procesamiento de datos.

Interacción con otros roles:

Proporciona datos limpios y estructurados a **Data Scientists** y **Data Business Analysts**.

Colabora con **Data Architects** para implementar y mantener la infraestructura de datos.

El proceso de *data science*: roles comunes, datos, obtención de datos, curación e ingeniería de características (*feature engineering*).

Data Business
analytics

- **Extracción de información:** Analizar datos para identificar *insights* que apoyen la toma de decisiones empresariales.
- **Reportes y métricas:** Crear informes y *dashboards* para monitorear el rendimiento del negocio.
- **Requisitos del negocio:** Traducir las necesidades empresariales en requerimientos técnicos para el equipo de datos.

Interacción con otros roles:

Trabaja con **Data Scientists** para obtener análisis avanzados.

Colabora con **Data Engineers** para asegurar que los datos necesarios estén disponibles.

Proporciona información clave a los **Usuarios**.

El proceso de *data science*: roles comunes, datos, obtención de datos, curación e ingeniería de características (*feature engineering*).

IT/Infraestructura
Engineer

- **Construcción de sistemas:** Diseñar y mantener la infraestructura tecnológica que soporta los sistemas de datos.
- **Seguridad y escalabilidad:** Asegurar que los sistemas sean seguros, escalables y eficientes.
- **Soporte técnico:** Resolver problemas de infraestructura y garantizar la disponibilidad de los servicios.

Interacción con otros roles:

Proporciona la infraestructura necesaria para **Data Architects** y **Data Engineers**.

Asegura que los sistemas cumplan con los requisitos de los **Usuarios**.

El proceso de *data science*: roles comunes, datos, obtención de datos, curación e ingeniería de características (*feature engineering*).

Usuario

- **Definición de objetivos:** Establecer las metas y necesidades del negocio que guían el trabajo del equipo de datos.
- **Retroalimentación:** Proporcionar comentarios sobre los resultados y ajustar los requisitos según sea necesario.
- **Toma de decisiones:** Utilizar los insights y análisis proporcionados por el equipo para tomar decisiones informadas.

Interacción con otros roles:

Trabaja estrechamente con **Data Business Analysts** para definir métricas y objetivos.

Recibe informes y visualizaciones de **Data Scientists** y **Data Business Analysts**.

Colabora con **IT/Infraestructura Engineers** para asegurar que los sistemas cumplan con las necesidades del negocio.

El proceso de *data science*: roles comunes, datos, obtención de datos, curación e ingeniería de características (*feature engineering*).

Actividad: "El proceso de los datos en funcionamiento"

Objetivo: Comprender los roles clave en un equipo de datos.

Instrucciones: en grupos, determinar los roles de cada uno a fin de proporcionarle a la Facultad un análisis del desempeño de los estudiantes de las distintas carreras para mejorar su retención y éxito académico.

¿Qué hace cada rol?

- *data scientist: análisis y visualización*
- *data architect: diseño y mantenimiento*
- *data engineer: disponibilidad e ingeniería*
- *data business analytics: extracción de información y reportes*
- *IT/infraestructura engineer: soporte técnico*
- *Usuario: toma de decisiones*

Big data: introducción y nociones básicas. Identificación de grandes volúmenes de datos

Big data, definido por Laney, un analista de Gartner (2001) es el conjunto de técnicas y tecnologías para el tratamiento de datos, en entornos de gran **volumen**, **variedad** de orígenes y en los que la **velocidad** de respuesta es crítica. Luego se añadió una cuarta V: **veracidad**.

La veracidad es un problema aún mayor para la analítica avanzada y la inteligencia artificial, donde el principio de «GIGO» (garbage in = garbage out), en cualquier sistema, la calidad de la salida está determinada por la calidad de la entrada.

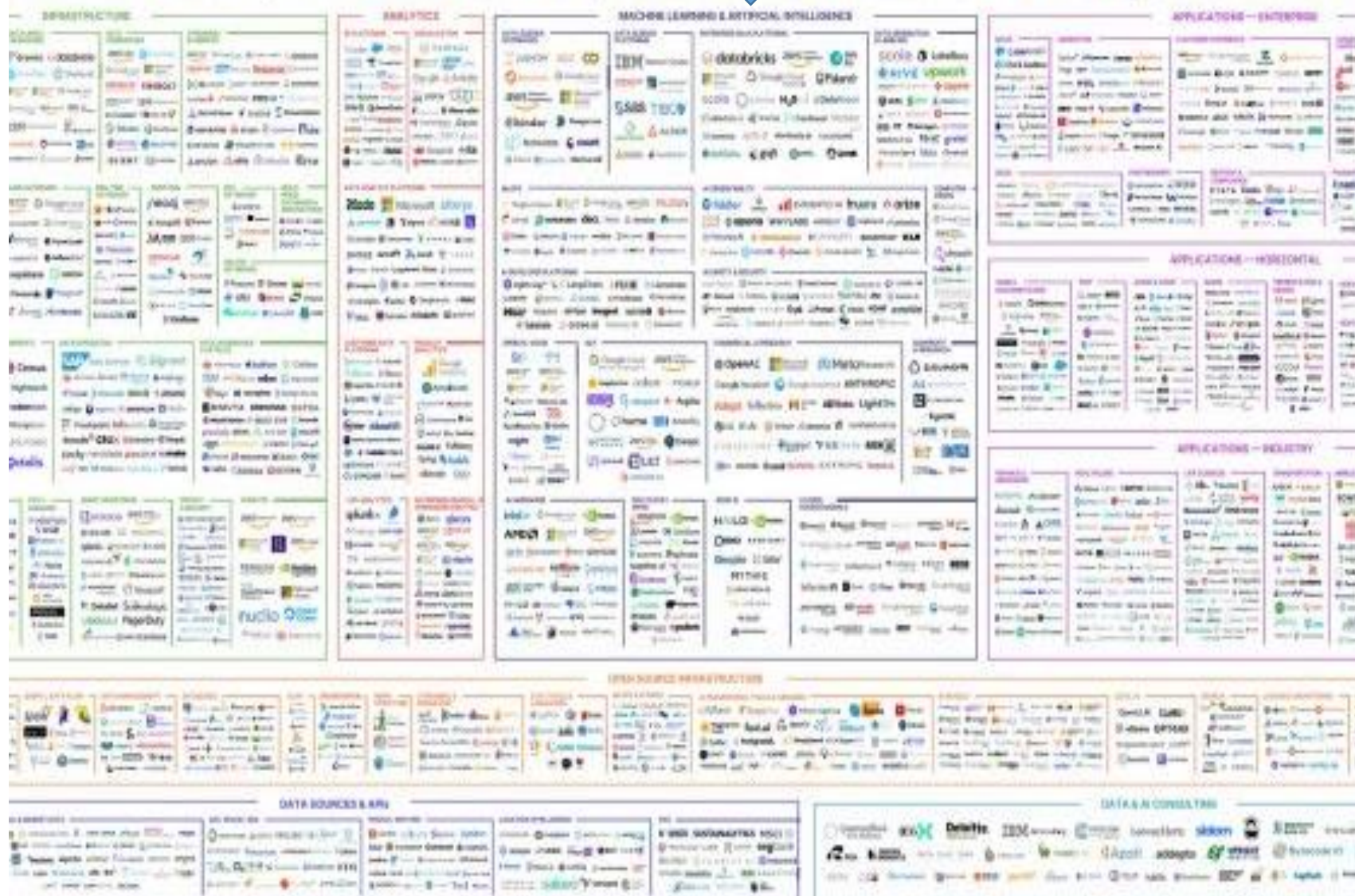
Big Data Landscape 2016



© Matt Turck (@mattturck), Jim Hao (@jimrhao), & FirstMark Capital (@firstmarkcap)

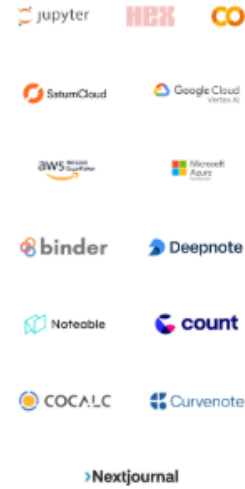
FIRSTMARK

THE 2024 MAD (MACHINE LEARNING, ARTIFICIAL INTELLIGENCE & DATA) LANDSCAPE



Machine Learning & Artificial Intelligence

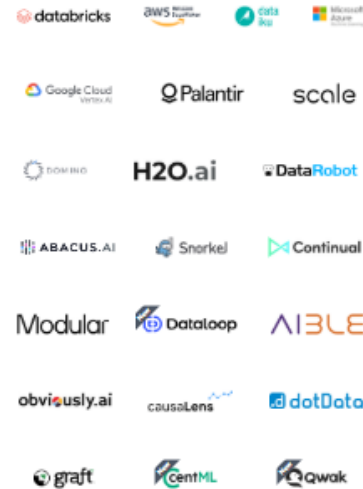
Data Science Notebooks



Data Science Platforms



Enterprise ML Platforms



Data Generation & Labelling



MLOps



AI Observability



AI Developer Platforms



AI Safety & Security



Arquitectura de datos

Arquitectura de Datos: Es un marco que define cómo los datos son recolectados, almacenados, transformados, distribuidos y consumidos en una organización. Incluye modelos, reglas, estándares y políticas que gobiernan qué datos se recopilan y cómo se utilizan.

Tipos de Arquitecturas de Datos

- **Arquitectura Tradicional:** Basada en data warehouses centralizados.
- **Arquitectura Moderna:** Incluye data lakes, arquitecturas basadas en la nube y soluciones híbridas.
- **Arquitectura Lambda y Kappa:** Diseñadas para el procesamiento de grandes volúmenes de datos en tiempo real y batch (conjunto de tareas que se procesan juntas en algún momento).

Arquitectura de datos

Componentes Clave

Data Sources (Fuentes de Datos): Los orígenes de los datos, como bases de datos, aplicaciones, sensores, etc.

Data Storage (Almacenamiento de Datos): Sistemas donde se almacenan los datos, como bases de datos relacionales, data lakes, data warehouses.

Data Processing (Procesamiento de Datos): Herramientas y tecnologías que transforman los datos en información útil.

Data Integration (Integración de Datos): Proceso de combinar datos de diferentes fuentes para proporcionar una vista unificada.

Data Governance (Gobierno de Datos): Conjunto de prácticas y políticas para asegurar la calidad, seguridad y cumplimiento de los datos.

Business Intelligence

Business Intelligence (BI) o **Inteligencia Empresarial** es un conjunto de herramientas, tecnologías y prácticas que permiten a las organizaciones transformar datos en información valiosa para la toma de decisiones. El objetivo principal del BI es ayudar a las empresas a analizar datos históricos y actuales, identificar tendencias y patrones, y generar insights que impulsen acciones estratégicas.

Componentes

- 1. Recopilación de Datos:** Obtener datos de diversas fuentes.
- 2. Almacenamiento de Datos:** Usar sistemas como Data Warehouses o Data Lakes para almacenar datos de manera centralizada.
- 3. Visualización de Datos:** Presentar los resultados de manera clara y comprensible mediante gráficos, dashboards e informes.
- 4. Análisis de Datos:** Aplicar técnicas y herramientas para explorar y analizar los datos.

Disponibilidad de datos

Es la capacidad de acceder y utilizar los datos cuando se necesitan. Esto implica que los datos deben estar:

Accesibles: Los usuarios autorizados deben poder encontrar y recuperar los datos fácilmente.

Fiables: Los datos deben ser precisos, completos y coherentes.

Oportunos: Los datos deben estar disponibles en el momento en que se necesitan para la toma de decisiones o para realizar tareas específicas.

Disponibilidad de datos

Importancia de la disponibilidad de datos

- **Toma de decisiones informadas:** Los datos disponibles permiten a las organizaciones tomar decisiones basadas en hechos y análisis, en lugar de intuiciones o suposiciones.
- **Operaciones empresariales fluidas:** Muchas operaciones empresariales dependen de la disponibilidad de datos para funcionar correctamente. Por ejemplo, los sistemas de gestión de inventario, los sistemas de atención al cliente y los sistemas de procesamiento de pagos requieren acceso constante a los datos.
- **Cumplimiento normativo:** En muchos sectores, las organizaciones están obligadas a mantener y proporcionar acceso a los datos para cumplir con las regulaciones y leyes aplicables.
- **Ventaja competitiva:** Las organizaciones que pueden acceder y utilizar los datos de manera eficaz pueden obtener una ventaja competitiva al identificar tendencias, anticipar necesidades y optimizar procesos.

Disponibilidad de datos

Factores que afectan la disponibilidad de datos

- **Infraestructura de TI:** La disponibilidad de datos depende de la fiabilidad y el rendimiento de la infraestructura de TI, incluidos los servidores, las redes y los sistemas de almacenamiento.
- **Seguridad de datos:** Las medidas de seguridad inadecuadas pueden provocar la pérdida o el robo de datos, lo que afecta su disponibilidad.
- **Gestión de datos:** Las prácticas de gestión de datos deficientes, como la falta de copias de seguridad o la falta de planes de recuperación ante desastres, pueden poner en peligro la disponibilidad de datos.
- **Errores humanos:** Los errores humanos, como la eliminación accidental de datos o la configuración incorrecta de los sistemas, también pueden afectar la disponibilidad de datos.

Estrategias para mejorar la disponibilidad de datos

- **Implementar soluciones de alta disponibilidad:** Estas soluciones garantizan que los datos estén disponibles incluso en caso de fallos del sistema o interrupciones.
- **Realizar copias de seguridad de datos con regularidad:** Las copias de seguridad permiten restaurar los datos en caso de pérdida o daño.
- **Implementar planes de recuperación ante desastres:** Estos planes describen los pasos que se deben seguir para restaurar los datos y los sistemas en caso de un desastre.
- **Establecer políticas y procedimientos de gestión de datos:** para que se gestionen de manera adecuada y segura.
- **Utilizar servicios de almacenamiento en la nube:** que pueden proporcionar alta disponibilidad y escalabilidad.

Gobernanza de datos

La gobernanza de datos empresariales es un marco de trabajo que define cómo una organización gestiona y utiliza sus datos, para asegurar que los datos sean de alta calidad, seguros, confiables y estén disponibles para las personas adecuadas en el momento adecuado.

La gobernanza de datos abarca **las políticas y procedimientos** que se implementan para garantizar que los **datos** de una organización sean **precisos y que se manejen correctamente** cuando se ingresan, almacenan, manejan, acceden y eliminan.

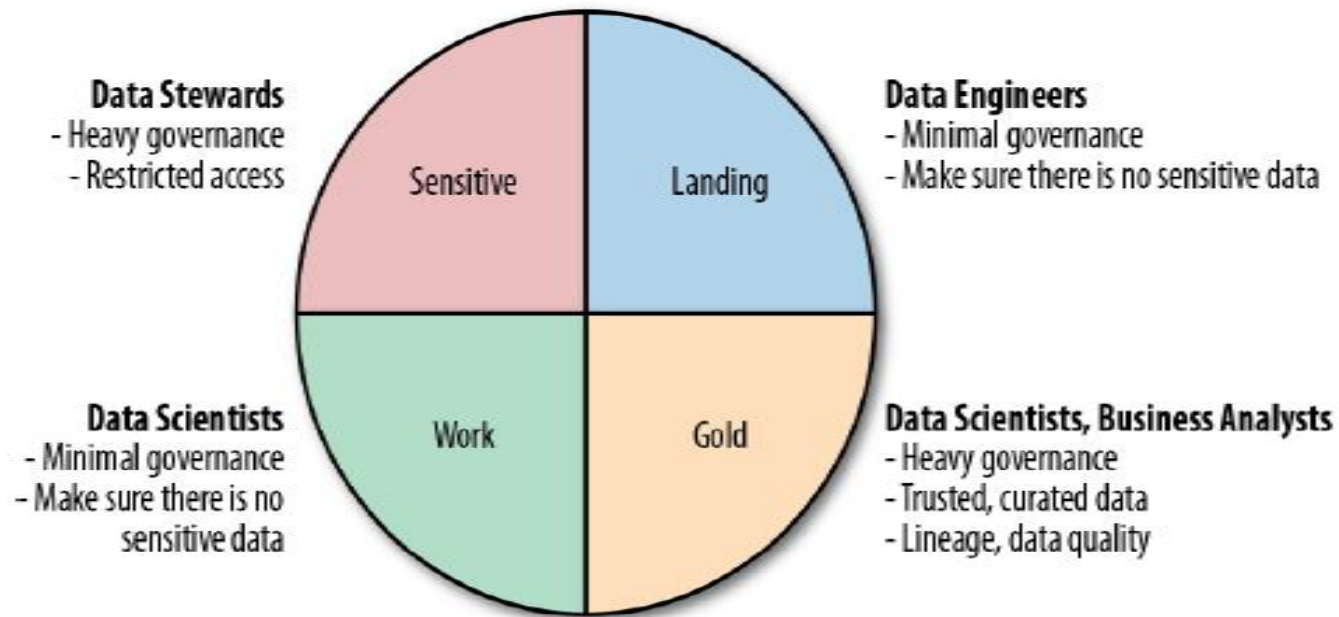
Las **responsabilidades** de gobernanza de datos incluyen establecer la infraestructura y tecnología, configurar y mantener procesos y políticas, e identificar a las personas (o cargos) de una organización que tienen la autoridad y responsabilidad de gestionar y salvaguardar tipos específicos de datos.

Permite aprovechar al máximo el valor de los datos. Ayuda a convertir los datos en un activo estratégico que impulsa el éxito empresarial

Gobernanza de datos

Un modelo de **gobernanza de datos** aplicado a un **lago de datos (data lake)**, segmentado en cuatro zonas distintas, cada una con diferentes expectativas de gobernanza y usuarios principales.

Un data lake es un repositorio centralizado que permite almacenar grandes volúmenes de datos en su formato nativo, ya sean estructurados, semiestructurados o no estructurados



- **Data Stewards:** Responsables de la calidad y gestión de datos en áreas específicas.
- **Data Engineers:** Responsables de la infraestructura y el flujo de datos.
- **Data Scientists:** Analistas que exploran y modelan los datos para obtener información valiosa.
- **Business Analysts:** Analistas que utilizan los datos para entender y mejorar los procesos de negocio.

Gobernanza de datos

- Sensitive (Sensible):

Usuarios: Data Stewards

Gobernanza: Alta (Heavy governance)

Acceso: Restringido (Restricted access)

Características: Contiene datos sensibles o confidenciales que requieren un alto nivel de protección y control.

- Landing (Aterrizaje):

Usuarios: Data Engineers

Gobernanza: Baja (Minimal governance)

Características: Zona donde los datos sin procesar "aterrizan" inicialmente. Se espera que no contenga datos sensibles.

- Work (Trabajo):

Usuarios: Data Scientists

Gobernanza: Baja (Minimal governance)

Características: Zona donde los Data Scientists exploran y transforman los datos para sus análisis. Se espera que no contenga datos sensibles.

- Gold (Oro):

Usuarios: Data Scientists y Business Analysts

Gobernanza: Alta (Heavy governance)

Características: Contiene datos de alta calidad, confiables y curados, listos para ser utilizados en análisis y reportes. Se mantiene el linaje de los datos y se asegura la calidad.

Ética en la ciencia de datos

Es un **sistema de derechos y responsabilidades para procesos basados en información**, que se ejecuta de acuerdo con modelos consensuados que describen quién puede tomar qué acciones, con qué información, y cuándo, bajo que circunstancias usando qué métodos. Data Governance Institute (2014).

Aspectos clave de la ética en los datos:

- **Privacidad:** Proteger la información personal de las personas y garantizar que se utilice de manera transparente y con su consentimiento. Cumplir con las leyes y regulaciones de privacidad de datos.
- **Sesgo:** Identificar y mitigar los sesgos en los datos y los algoritmos para evitar la discriminación y la injusticia.
- **Transparencia:** Explicar cómo se recopilan, utilizan y comparten los datos. Hacer que los algoritmos y los modelos sean comprensibles y explicables.
- **Responsabilidad:** Asumir la responsabilidad por el impacto de los datos y los algoritmos. Establecer mecanismos de rendición de cuentas para garantizar que los datos se utilicen de manera ética.
- **Seguridad:** Proteger los datos contra accesos no autorizados, pérdidas o robos. Garantizar la integridad y la confidencialidad de los datos.

Ética en la ciencia de datos

Importancia de la ética en los datos:

•Generar confianza:

- El uso ético de los datos genera confianza entre las organizaciones y sus clientes.
- La confianza es esencial para el éxito a largo plazo de cualquier organización.

•Evitar daños:

- El uso indebido de los datos puede causar daños a las personas y a la sociedad.
- La ética en los datos ayuda a prevenir estos daños.

•Promover la justicia:

- El uso ético de los datos puede ayudar a promover la justicia y la igualdad.
- Los datos pueden utilizarse para identificar y abordar las desigualdades.

•Cumplir con las regulaciones:

- Muchas jurisdicciones han implementado leyes y regulaciones de privacidad de datos.
- La ética en los datos ayuda a las organizaciones a cumplir con estas regulaciones

Ética en la ciencia de datos

La ética en los datos se refiere al conjunto de principios y valores que guían la recopilación, el uso y la gestión de la información.

Actividad: "El Uso Ético de los Datos en un Estudio de Comportamiento del Consumidor"

Objetivo: Evaluar críticamente las implicaciones éticas de las prácticas de recolección, uso y almacenamiento de datos en un estudio de comportamiento del consumidor, identificando posibles conflictos y proponiendo soluciones éticamente sólidas.

Instrucciones: en grupos, a partir de un caso de estudio reflexionar sobre algunas prácticas.

Ética en la ciencia de datos

Caso de Estudio: El Uso Ético de los Datos en un Estudio de Comportamiento del Consumidor

Una empresa de comercio electrónico quiere realizar un estudio de comportamiento del consumidor para mejorar sus campañas de marketing. Contratan a un equipo de científicos de datos para analizar una gran cantidad de información, incluyendo:

- Datos personales de los clientes (nombres, direcciones, edad, género).
- Historial de compras (productos comprados, fechas, montos pagados).
- Actividad en el sitio web (páginas visitadas, tiempo en cada página, clics).
- Información obtenida a partir de cookies y otras herramientas de seguimiento (ubicación aproximada, historial de navegación fuera del sitio web).

El objetivo del análisis es identificar patrones de compra y desarrollar un sistema de recomendaciones personalizadas para aumentar las ventas.

Durante el análisis, el equipo se da cuenta de que la empresa no ha informado claramente a los usuarios sobre el uso de sus datos más allá del propósito del procesamiento de sus compras (por ejemplo, para personalizar futuras campañas de marketing). Además, los clientes no dieron su consentimiento explícito para que la empresa recopile datos sobre su comportamiento en otros sitios web.

Ética en la ciencia de datos

Caso de Estudio: El Uso Ético de los Datos en un Estudio de Comportamiento del Consumidor

Preguntas para reflexionar:

1. ¿Qué dilemas éticos presenta este caso?

- ¿Es ético usar todos los datos de los clientes si no dieron su consentimiento explícito para ciertos usos?
- ¿Es correcto usar herramientas de seguimiento (como cookies) para obtener datos de usuarios fuera del sitio sin informarles claramente?
- ¿Qué problemas pueden surgir si la empresa comparte esta información con terceros?

2. ¿Cómo debería proceder el equipo de científicos de datos para resolver este dilema ético?

- ¿Deben continuar con el análisis utilizando todos los datos disponibles?
- ¿Qué medidas de transparencia y consentimiento deberían tomar antes de proseguir?
- ¿Cómo deben garantizar la privacidad y el respeto a los derechos de los clientes?

Ética en la ciencia de datos

Caso de Estudio: El Uso Ético de los Datos en un Estudio de Comportamiento del Consumidor

Solución del caso:

1. Identificación de dilemas ético:

- **Transparencia y consentimiento:** Los clientes no han sido informados de manera clara y explícita sobre el uso de sus datos para fines de marketing personalizado o sobre la recolección de información fuera del sitio web. El uso de cookies sin un consentimiento adecuado puede violar las leyes de protección de datos.
- **Uso justo de los datos:** Aunque los datos recopilados pueden mejorar las ventas y la experiencia del cliente, su uso sin el consentimiento adecuado plantea preocupaciones éticas sobre la violación de la privacidad.
- **Riesgo de compartir datos con terceros:** Si los datos se comparten con otras empresas o plataformas sin el conocimiento o consentimiento de los usuarios, se podrían producir abusos, como la venta o mal uso de la información personal.

Ética en la ciencia de datos

Caso de Estudio: El Uso Ético de los Datos en un Estudio de Comportamiento del Consumidor

Solución del caso:

2. Pasos éticos a seguir:

- **Revisar la política de privacidad y consentimiento:** El equipo de científicos de datos debe recomendar a la empresa que actualice sus políticas de privacidad para garantizar que los usuarios sean informados de manera clara y precisa sobre cómo se recopilarán, almacenarán y utilizarán sus datos. Esto incluiría una explicación explícita del uso de cookies y de la recopilación de información fuera del sitio web.
- **Obtener consentimiento informado:** Los usuarios deben dar su consentimiento explícito para el uso de sus datos con fines de marketing y para la recolección de datos sobre su comportamiento en otros sitios web. El equipo de científicos de datos debe garantizar que no se utilicen datos sin este consentimiento.
- **Minimización de datos:** En lugar de utilizar todos los datos disponibles, el equipo puede optar por **minimizar el uso de información sensible**, es decir, recolectar y analizar solo los datos necesarios para el análisis, y asegurarse de que los datos personales estén anonimizados siempre que sea posible.
- **Cumplimiento normativo:** El equipo debe garantizar que sus acciones cumplan con las leyes locales e internacionales de protección de datos.

Ética en la ciencia de datos

Caso de Estudio: El Uso Ético de los Datos en un Estudio de Comportamiento del Consumidor

Solución del caso:

2. Pasos éticos a seguir:

- **Transparencia y comunicación con los usuarios:** La empresa debe comunicar a sus usuarios los beneficios de proporcionar sus datos, como la posibilidad de recibir recomendaciones personalizadas, pero sin imponer esta opción. Los usuarios deben tener la **posibilidad de optar por no participar** si así lo desean.
- **Transparencia y consentimiento:** Los clientes no han sido informados de manera clara y explícita sobre el uso de sus datos para fines de marketing personalizado o sobre la recolección de información fuera del sitio web. El uso de cookies sin un consentimiento adecuado puede violar las leyes de protección de datos.
- **Uso justo de los datos:** Aunque los datos recopilados pueden mejorar las ventas y la experiencia del cliente, su uso sin el consentimiento adecuado plantea preocupaciones éticas sobre la violación de la privacidad.
- **Riesgo de compartir datos con terceros:** Si los datos se comparten con otras empresas o plataformas sin el conocimiento o consentimiento de los usuarios, se podrían producir abusos, como la venta o mal uso de la información personal.

Ética en la ciencia de datos

Caso de Estudio: El Uso Ético de los Datos en un Estudio de Comportamiento del Consumidor

Solución del caso:

El equipo de científicos de datos no debería continuar el análisis con todos los datos sin antes garantizar que se han respetado los principios éticos fundamentales, como:

- el consentimiento informado,
- la minimización de datos
- la protección de la privacidad.

Además, la empresa debe comprometerse con la transparencia y ajustar sus políticas de datos para respetar los derechos de los clientes.

Software de programación en ciencia de datos

- El software de programación en ciencia de datos se refiere a las herramientas y plataformas que permiten a los científicos de datos realizar tareas como recopilación y limpieza de datos, análisis exploratorio de datos (EDA), modelado y aprendizaje automático, visualización de datos

Importancia del Software de Programación en Ciencia de Datos:

- Permite automatizar tareas repetitivas y complejas.
- Facilita la exploración y el análisis de grandes volúmenes de datos.
- Permite construir modelos predictivos precisos.
- Mejora la comunicación de los resultados a través de visualizaciones interactivas.