

Etapa 1: Fundamentos y Calidad de Datos (Clases 1 a 4)

Objetivo: Comprender los conceptos básicos de BI y evaluar la calidad de los datos de ventas. Actividades:

1. Captación del cliente: Diseñar un breve informe (en Google Docs) que explique la importancia de la aplicación de BI para un e-commerce. Incluir ejemplos de herramientas de BI, cómo las va a utilizar y los beneficios para el negocio.

En la actualidad, es importante analizar los datos para transformarlos en nuevos conocimientos, los cuales nos permitirán tomar decisiones tanto estratégicas como operativas. Business Intelligence es un conjunto de tecnologías, aplicaciones y prácticas que permitirán recopilar, analizar y visualizar los datos para facilitar la toma de decisiones. En una e-commerce, BI mejorará la toma de decisiones, permitirá reaccionar rápidamente a las tendencias del mercado, aumentar las ventas, mejorar la experiencia del cliente, optimizar las oportunidades y mejorar en la empresa.

Utilizar BI permite utilizar diversas herramienta para la recopilación de los datos:

- Data Warehousing
 - Se utilizan para recopilar y almacenar grandes volúmenes de datos que provienen de diversas fuentes
- Herramientas de analítica:
 - Estas herramientas permiten el análisis avanzado de los datos
 - Algunos softwares utilizados son:
 - Power BI
 - Tableau
 - Qlik Sense
- ETL (Extract, Transform, Load)
 - Son procesos y herramientas que permiten extraer datos de diversas fuentes y transformarlos en un formato para luego cargarlos en el sistema de almacenamiento de datos
- Visualización de datos:
 - Permiten representar los datos en forma gráfica, facilitando su entendimiento y el análisis de los mismos

2. Investigación (contexto ético y legal): Realizar un caso de estudio sobre una violación de ética en datos en el ámbito del comercio electrónico nacional o internacional. Elaborar un informe analizando el contexto, las faltas y las consecuencias, así como las lecciones aprendidas y errores a evitar.

Empresa internacional de comercio electrónico, que realiza operaciones en Argentina, lanzó una campaña para mejorar sus estrategias de ventas enfocadas en consumidores millennials. Para esto, comenzaron una investigación de mercado cuyo objetivo era identificar cuáles eran los productos en tendencia en ese segmento, a fin de anticipar la demanda y abastecer su tienda online.

La estrategia de investigación consistió en recolectar grandes volúmenes de datos sobre comportamiento de compra, navegación web y actividad en redes sociales. Si bien parte de esta información era obtenida a través de sus propios canales (como búsquedas internas y compras previas), también contrataron a una empresa externa que ofrecía acceso a información adicional, como historiales de búsqueda en sitios de terceros, publicaciones en redes sociales y datos demográficos.

Esta información fue utilizada para alimentar algoritmos de inteligencia artificial que segmentaban a los usuarios y predecían qué productos serían más populares entre los millennials. A partir de allí, se tomaron decisiones sobre qué productos importar y ofrecer en el portal.

Faltas éticas y legales identificadas

- **Recolección sin consentimiento adecuado:** muchos de los datos utilizados provenían de cookies de terceros y técnicas de rastreo sin que los usuarios tuvieran una información clara y específica sobre la recolección. En algunos países donde operaba la empresa, esto violaba normativas como el Reglamento General de Protección de Datos (GDPR) de la Unión Europea y leyes locales de protección de datos.
- **Uso de datos sensibles y perfilado:** se construyeron perfiles detallados de usuarios millennials, incluyendo aspectos como preferencias personales, hábitos de consumo y estado económico. Esto se realizó sin mecanismos adecuados de anonimización ni salvaguardas éticas para proteger a los usuarios frente a un tratamiento invasivo de sus datos.
- **Falta de transparencia:** la empresa no comunicó de manera clara a sus clientes que se estaban utilizando datos de múltiples fuentes para hacer predicciones automatizadas sobre su comportamiento.
- **Tercerización de datos sin control:** al contratar a una empresa externa para obtener datos, no se garantizó que esa empresa también respetara las normativas de protección de datos ni se auditaron sus prácticas.

Consecuencias

- **Sanciones económicas:** en algunos países donde la empresa operaba, se iniciaron investigaciones por violaciones a la legislación sobre privacidad. En la Unión Europea, por ejemplo, se le impuso una multa por infringir el principio de transparencia y consentimiento del GDPR.
- **Pérdida de confianza:** la noticia fue cubierta por varios medios de comunicación y generó una reacción negativa entre los usuarios, que sintieron que su privacidad había sido vulnerada. Muchos decidieron dejar de utilizar la plataforma.

- Reputación corporativa: además del impacto económico, la marca sufrió una crisis reputacional, especialmente entre los propios millennials, su público objetivo, quienes son más sensibles a temas de privacidad digital.

Lecciones aprendidas

- Consentimiento informado: toda recolección de datos debe estar precedida por un consentimiento claro, específico, informado y revocable por parte del usuario.
- Transparencia en el tratamiento de datos: los usuarios deben saber para qué se recopilan sus datos y cómo se van a usar. Esto incluye explicar si serán utilizados por algoritmos de inteligencia artificial.
- Ética en la segmentación: aunque el perfilado puede mejorar la eficiencia comercial, se debe evitar cruzar la línea entre el análisis legítimo y la manipulación o invasión de la privacidad.
- Responsabilidad extendida: las empresas deben asegurarse de que sus socios y proveedores también respeten las normativas éticas y legales en el tratamiento de datos.

Errores a evitar en el futuro

- Utilizar datos sin verificar la fuente y el consentimiento original.
- Realizar tratamientos automatizados sin mecanismos de revisión humana o sin evaluar su impacto en la privacidad.
- Asumir que el marketing predictivo siempre está justificado sin evaluar sus implicancias sociales o personales.
- Tercerizar el análisis de datos sin cláusulas claras de responsabilidad y auditoría sobre el uso de la información

3. Investigación (exploración y búsqueda de fuentes de datos): Identificar tres fuentes públicas de datos relacionadas con el e-commerce, como estadísticas sobre ventas en línea o datos de comportamiento del consumidor. Crear una tabla que resuma cada fuente con enlaces, tipos de datos y potencial de análisis.

Fuente	Enlace	Tipo de datos	Potencial de análisis
Cámara Argentina de Comercio Electrónico (CACE)	https://cace.org.ar/	Encuestas a usuarios y empresas, facturación, órdenes de compra, ticket promedio	Analizar tendencias de consumo, segmentación de usuarios, evolución de ventas y comportamiento del comprador
Statista – E-commerce en Argentina	https://es.statista.com/temas/11615/e-commerce-en-argentina/	Estadísticas de ventas, usuarios, plataformas, penetración del e-commerce	Comparar mercados, identificar productos populares, analizar crecimiento y hábitos de compra
Wuala.net – Datos Clave de E-commerce	https://wuala.net/datos-clave-de-2024-y-lo-que-espera-en-2025/	Facturación, ticket promedio, categorías destacadas, comportamiento del comprador	Entender el perfil del consumidor, identificar categorías líderes y tendencias de g

4. Obtención de datos: Importar el conjunto de datos de ventas a Google Sheets. Crear un formulario con el checklist de criterios de calidad de datos y aplicar el checklist al conjunto de datos, registrando los hallazgos en un informe breve.

En la revisión de los datos de ventas del e-commerce, se encontraron con algunos detalles importantes que vale la pena mencionar. Se encontraron valores faltantes en ID_Ventas, esta casilla no debería aceptar valores nulos, ya que algunos registros no están completamente identificados. También se detectaron datos duplicados, por lo que se procede a eliminar dichas duplicaciones.

Con respecto al formato de las fechas no es uniforme: algunas aparecen con un solo dígito para el día (por ejemplo, 4/12/2024) y otras con dos (04/12/2024). Esto puede complicar la lectura.

Los datos no están completamente actualizados, están hasta el periodo 2024.

Se recomienda:

- Eliminar los duplicados.
- Unificar el formato de fechas.
- Actualizar los datos.

Form: https://docs.google.com/forms/d/e/1FAIpQLSeT7-fvMgu_nwdxatzkWVxGMkHkhU5LnJkArrD1UULCbialsQ/viewform?usp=dialog