

Clase

¿Qué es ciencia de datos y big data? ¿Cómo afectan a mi negocio?

Reto

Piensa en 3 empresas que almacenan un gigantesco volumen de datos y están usando **big data** hoy en día.

¿Qué tipo de información podemos analizar?

Piensa qué **tipos de datos** manejan las 3 empresas del reto anterior y para qué crees que los están utilizando.

¿Cómo crear empresas y culturas data-driven?

Piensa en 3 empresas que tengan una **cultura de datos** (*data-driven*), donde todas sus decisiones se toman con base en la información que generan.

¿Qué es inteligencia artificial y machine learning?

Piensa en 3 empresas (*de 3 sectores/industrias diferentes*) que están utilizando **lenguaje automatizado**.

¿Qué es deep learning?

Piensa en 3 empresas que producen tecnología de **deep learning** y para qué crees que la están utilizando (*1 con imágenes, 1 con video y 1 con sonido*).

Flujo de trabajo en ciencia de datos: fases, roles y oportunidades laborales

¿Cuál es tu rol? Piensa en dónde te proyectas, qué área(s) de la ciencia de datos te llaman la atención, cuáles herramientas te llamaron la atención y qué propósito quieres cumplir como data scientist.

Herramientas para cada etapa del análisis de datos

Acaban de informarte que las ventas de tu empresa cayeron mucho hoy. Debes **extraer** la información de los clientes que compraron ayer y los clientes que compraron hoy para entender por qué cayeron las ventas. **¿Qué herramienta debes usar para realizar la extracción de datos?**

¿Qué es y cómo usar una base de datos relacional con SQL?

Vas a analizar las ventas registradas en tu base de datos, pero debes **excluir** del análisis cualquier registro donde la columna *city* sea *Bogotá*. **¿Cuál operador lógico necesitas para obtener la información con este "filtro"?**

Cómo estructurar queries en SQL

En la pestaña **VENTAS_2020** observa la información de la "base de datos", plantéate una pregunta y encuentra una solución por medio de queries en SQL.

Conflictos y retos actuales sobre la ética y tratamiento de datos

Mira el documental *The Social Dilemma*, piensa qué crees que es verdad o no y luego mira el siguiente análisis:

https://www.youtube.com/watch?v=Frdna_cBwwU.

¿Qué aprendiste? ¿Acertaste en algunos de tus puntos? ¿Qué fue lo que más te asustó o sorprendió?

Aplica técnicas de storytelling para convertir problemas de datos en historias

Piensa en un **problema de negocio** en la empresa donde trabajas (*o donde te gustaría trabajar*) y aplica la **estructura de storytelling**:

- ¿Cómo lo identificas?
- ¿Qué solución podría tener?
- ¿Cuál es su alcance?

Cómo estructurar un caso de negocio

Valida la hipótesis que planteaste en el reto anterior, encuentra sus **variables cuantitativas y cualitativas**.

Análisis cuantitativo en un caso de negocio

¿Cómo sería tu **clasificación numérica** para resolver el caso de negocio de los retos anteriores?
- ¿Qué usuarios buscarías?
- ¿Cuál sería tu hipótesis?

Recuerda que no necesitas llegar a un número, solo definir tu **estrategia e hipótesis numérica**.

Análisis cualitativo en un caso de negocio

Identifica y **clasifica las variables categóricas** del caso de negocio que has estado trabajando en los retos anteriores.

Fusión cuanti-cualitativa en un caso de negocio

Piensa cómo poner en conjunto la **información cuantitativa y cualitativa** en tu caso de negocio.

¿Qué es minería de texto? ¿Cómo usarla para obtener información adicional?

Piensa en los **mensajes** que podría haber dentro de las categorías que creaste. ¿Qué **información adicional** podrías encontrar **analizando los textos** relacionados a tu caso de negocio?
¿Cuáles serían los diferentes comportamientos?

Variación de comportamientos a partir de la geolocalización

Plantea una hipótesis de **cómo afecta la geolocalización** al comportamiento de tus usuarios y define el campo geográfico que más se adecua a tu caso de estudio (*ciudades, pueblos, distritos, países...*).

Acciones, algoritmos y toma de decisiones según los resultados del análisis

Piensa qué acciones surgirían como resultado de tu análisis.

Continúa aprendiendo ciencia y análisis de datos para ejecutar estrategias efectivas

¡Felicidades por terminar el Curso de Análisis de Negocios para Ciencia de Datos!
¿Aprendiste algo nuevo? ¿Qué fue lo que más te gustó? ¿Cuál será tu siguiente paso para convertirte en data scientist?

Solución

1. Google
2. Microsoft
3. Apple

****Tipos de datos que recojen las empresas****

1. Google:

- Personas: "from figuring out basic stuff like which language you speak."
- Transacciones: "if you add a credit card or other payment method to you
- Navegación web: "We collect information about your activity in our serv
- Machine 2 Machine: "we store the information we collect with unique ic

2. Microsoft:

- Personas: "Name and contact data. Your first and last name, email addre
- Transacciones: "Payment data. Data to process payments, such as your payment instrument. "
- Navegación web: "Interactions. Data about your use of Microsoft produ
- Machine 2 machine: "Data about your device, your device configuratio device, including product keys. In addition, IP address, device identifiers (such

3. Apple:

- Personas: ""Contact Information. Data such as name, email address, phy
- Transacciones: "Payment Information. Data about your billing address a
- Navegación web: "Usage Data. Data about your activity on and use of o interaction; crash data, performance and other diagnostic data; and other usa
- Machine 2 machine: "Device Information. Data from which your device
- Biométricos: "Face ID data — including mathematical representations o

****Fuentes:****

1. Google:

- <https://policies.google.com/privacy?hl=en-US>

2. Microsoft:

- <https://privacy.microsoft.com/en-us/privacystatement>

3. Apple:

- <https://www.apple.com/legal/privacy/en-ww/>
- <https://www.apple.com/legal/privacy/data/en/face-id/>

1. Netflix

2. Facebook

3. Youtube

1. Procter & Gamble: "Utiliza machine learning para optimizar su cadena de su logística, reducir costos y asegurarse de que los productos estén disponibles e

2. American Express: "Emplea machine learning para detectar transacciones fi identificar posibles actividades sospechosas y proteger las cuentas de sus usu

3. Walmart: "Utiliza machine learning para optimizar la gestión de inventario e como el clima, festividades y tendencias de compra, asegurando que tengan s

1. Nestlé (Imágenes): Nestlé utiliza el deep learning en imágenes para mejorar la empresa puede identificar defectos o anomalías en los productos, aseguran
2. Disney (Video): Disney utiliza el deep learning en el ámbito de la animación la generación de personajes y escenas, la compañía puede simular comportam las películas de Disney ofrezcan experiencias visuales asombrosas y sumerjan
3. Spotify (Sonido): Spotify utiliza deep learning en el procesamiento de señal patrones complejos en las canciones que escuchan los usuarios, el algoritmo c del cliente y el tiempo de reproducción en la plataforma.

Quiero ser... Data Scientist

Debo usar... SQL

Necesito el operador lógico... NOT

- ¿Cuál fue tu pregunta?
Cual es el total de las ventas desde inicio del siglo XXI;

- ¿Cuál fue tu query en SQL?
SELECT SUM (VALOR)
FROM VENTAS_2020
WHERE AÑO >= 2000

- ¿Qué solución/respuesta encontraste?
La suma de todas las ventas desde principios de siglo XXI es 2285\$ MXN

- Aprendí que ... Los sistemas de recomendación no son pensantes,
- Acerté en ... La forma en la que los sistemas están diseñados nos ayuda a entenderlos mejor.
- Lo que más me sorprendió fue ... Los grafos y cómo funcionan los algoritmos de recomendación.

- Identifiqué ... Que existe una marca de uniformes de karate (karate gis en japonés) que es muy complicada, más si es para Latinoamérica, la página web está en japonés y la solución puede ser ... Ofrecer traducción de la página a inglés y español, mejor para reducir los costos y tiempos de envío.
- Su alcance es ... Toda Latinoamérica, que es una de las regiones más apartadas del mundo.

- Mi hipótesis fue... La empresa de uniformes de karate no vende tantos productos como podría.

- Variables cuantitativas: ...
 1. Relación de visitas de la página con el número de compras hechas por la página
 2. Tiempo de envío de los productos
 3. Tiempo de confirmación del pedido del producto
 4. Relación de la cantidad de veces que un producto se agrega al carrito pero no se compra

- Variables cualitativas: ...
 1. Evaluación del diseño de la página
 2. Análisis de comentarios en foros
 3. Validación de los procesos de manufacturación
 4. Comparaciones con la competencia

- Variables cuantitativas:

1. Relación de visitas de la pagina con el numero de compras hechas por la pagina
2. Tiempo de envio de los productos
3. Tiempo de confirmación del pedido del producto
4. Relación de la cantidad de veces que un producto se agrega al carrito pero no se

1. Descargar información:

¿Cuales pueden ser las mayores causas por las que nuestras ventas no son tan altas?
-Relación de visitas de la pagina con el numero de compras hechas por la pagina

2. Identificación:

Para identificar las posibles causas tenemos que crear hipotesis de variables que nos den pistas, como podrían ser, las cuales serían:

1. Conocimiento del idioma en el cual está la pagina
2. Lugares de residencia de los clientes
3. Tiempo de manufacturación de los productos
4. Cantidad de pagos procesados

3. Definir:

Definimos los clientes y los separamos por región de residencia, idioma y lo compra

Métrica	Norteamérica	Europa	Asia	Latinoamérica	Centroamérica
Visitas	1200	800	1500	500	600
Compras	50	30	70	10	30
Ratio de Compras por Visitas	4.17%	3.75%	4.67%	2.00%	5.00%

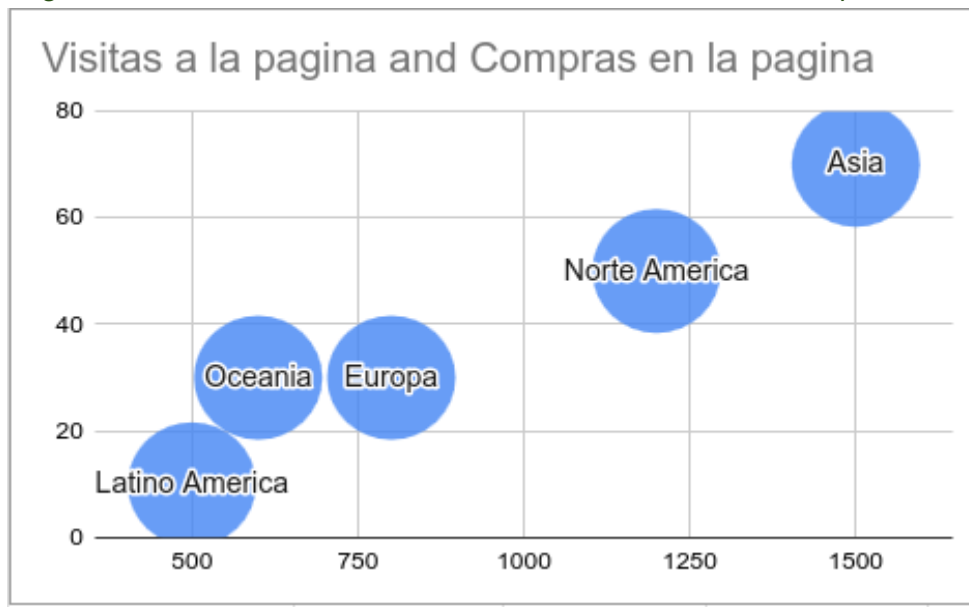
- Variables cualitativas: ...

1. Suboptimo diseño de la pagina 35%
2. Tiempos de los envios 25%
3. Tiempos de los procesos de manufacturación 20%
4. Ventajas de la competencia sobre la empresa 20%

	Suboptimo diseño de la pagina	Tiempos de los envios	Tiempos de los procesos de manufacturación	Ventajas de la competencia sobre la empresa
Norte America	30%	25%	20%	25%
Europa	35%	20%	25%	20%
Asia	25%	30%	20%	25%
Latino America	40%	15%	20%	25%
Oceania	20%	30%	25%	25%

- El mejor uso que podría darle a la minería de texto en mi caso de negocio es ... Arreglar la variable cualitativa de las ventajas de la competencia sobre nuestra empresa

- La geolocalización de mis usuarios afecta en... La cantidad de visitas y la cantidad de compras



- Acciones:

1. separar clientes por región
2. Contactar usuarios de las regiones con menos ventas
3. Aplica la retroalimentación de los usuarios
4. Validar

- Lo que más me gustó del curso fue... Que nos adentramos realmente a hacer
- Lo que más me sorprendió fue... Todo lo que hay dentro del proceso de análisis
- Mi siguiente paso para convertirme en data scientist será... Nunca parar de aprender

your Google Account, you can use it to buy things across our services, like apps in the Play Store." "We use these services, which we use to do things like recommend a YouTube video you might like." "Identifiers tied to the browser, application, or device you're using."

"Email address, postal address, phone number, and other similar contact data." "Payment instrument number (such as a credit card number) and the security code associated with your"

cards. In some cases, such as search queries." "Location, and nearby networks. For example, data about the operating systems and other software installed on your device, such as the IMEI number for phones)."

"Physical address, phone number, or other contact information" "Name and method of payment, such as bank details, credit, debit, or other payment card information" "Your interactions with our offerings, such as app launches within our services, including browsing history; search history; product usage data" "Information that could be identified, such as device serial number, or about your device, such as browser type" "Information about your face — is encrypted and protected with a key available only to the Secure Enclave."

"The minister can predict the demand for products in different regions. This allows them to improve efficiency in the moment and in the right place."

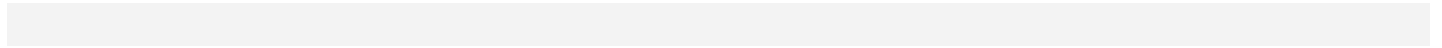
"Fraudulent in real time. The system analyzes spending patterns and customer behavior for fraud." "Fraudulent in real time. The system analyzes spending patterns and customer behavior for fraud."

"In their stores. With advanced algorithms, they can predict the demand for products based on factors such as sufficient stock available and avoiding excess."

la calidad del control de alimentos y bebidas en su proceso de producción. Mediante el análisis de imágenes, do que solo los artículos de alta calidad lleguen al mercado.

y efectos visuales para crear películas más realistas y cautivadoras. Al emplear algoritmos de deep learning en ientos y movimientos más naturales, así como mejorar la calidad de los efectos especiales. Esto permite que aún más al público en sus historias mágicas.

es de audio para mejorar su sistema de recomendación de música. Mediante el análisis de características y le deep learning puede ofrecer recomendaciones más precisas y personalizadas, aumentando la satisfacción



*son algoritmos de inteligencia artificial
hacen mas extremistas
s sistemas de recomendación.*

*l que se llama shureido, es una marca de renombre a nivel mundial, el problema es que poder comprar sus productos
los tiempos para preparar el uniforme y enviarlo a su destino toma meses.
ar en general la UX y UI da la pagina, y abrir fabricas en latino america o realizar convenios con otras tiendas locales
e esta marca.*

e compra

¿ compra

ltas como podrían? Para ello vamos a descargar la siguiente información:

os ayuden a responder a la pregunta "¿Cuales pueden ser las mayores causas por las quenuestras ventas no son tan

ramos con Relación de visitas de la pagina con el numero de compras hechas por la pagina

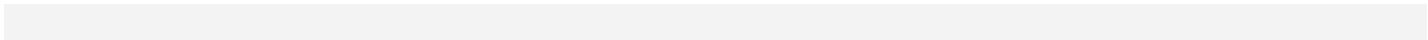
Oceanía

30
10
30%

e los le ración	Ventajas de la competencia sobre la empresa
20%	25%
25%	20%
20%	25%
20%	25%
25%	25%

alizando los comentarios en redes sociales los que le daría a una compañía una visión más amplia y desagregada esa

d de compras por la pagina



r analisis
isis
aprender

Día	Mes	Año	Producto	ID	Valor
1	2	1998	Bocina	24	\$528
12	4	2004	Auriculares	31	\$240
14	8	2016	Auriculares	14	\$315
16	10	2019	Bocina	200	\$105
21	12	2020	Bocina	304	\$680