



Programa PUCP DATA LITERACY

Formación tecnológica

Dirección Académica de Planeamiento y Evaluación
Dirección de Gestión del Talento Humano





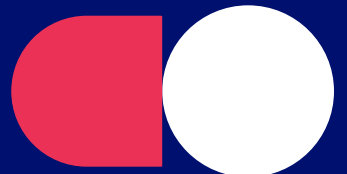
Contenido:



- 1 Bienvenida al curso
- 2 ¿Qué es Data Literacy?
- 3 ¿Qué involucra?
- 4 Pirámide DIKW



Bienvenida al curso





Daniela Estrada Nates

- CEO & Cofundadora de Danalytics
- Más de 12 años de experiencia en proyectos de analítica
- Profesora de maestría de Ciencia de Datos, MBA y Especialización en Analítica de Datos de la Universidad Icesi (2017 - 2021)
- Coordinadora académica de la especialización en analítica de Datos de la Universidad Icesi (2018 - 2020)
- Certificada en Data Literacy por Data Literacy LLC, compañía especializada en entrenamientos de Data Literacy

Temas	Tema	Contenido	Tiempo
1	Introducción a Data Literacy	- Definición e impacto de Data Literacy - Pirámide DIKW	2 horas

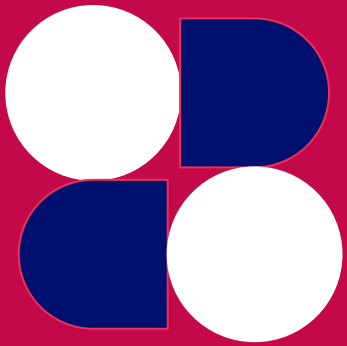
Temas	Tema	Contenido	Tiempo
1	Introducción a Data Literacy	- Definición e impacto de Data Literacy - Pirámide DIKW	2 horas
2	Tipos de datos y limpieza de Datos	- Tipos de datos y niveles de medición - Limpieza básica desde Power Query	6 horas

Temas	Tema	Contenido	Tiempo
1	Introducción a Data Literacy	- Definición e impacto de Data Literacy - Pirámide DIKW	2 horas
2	Tipos de datos y limpieza de Datos	- Tipos de datos y niveles de medición - Limpieza básica desde Power Query	6 horas
3	Análisis de Datos Descriptivo y Gráficas	- Datos en forma de valores absolutos, porcentajes, razones, variaciones porcentuales - Datos como medidas de tendencia central, dispersión y posición - Selección de gráficas básicas adecuadas para la representación visual de datos	10 horas

Temas	Tema	Contenido	Tiempo
1	Introducción a Data Literacy	- Definición e impacto de Data Literacy - Pirámide DIKW	2 horas
2	Tipos de datos y limpieza de Datos	- Tipos de datos y niveles de medición - Limpieza básica desde Power Query	6 horas
3	Análisis de Datos Descriptivo y Gráficas	- Datos en forma de valores absolutos, porcentajes, razones, variaciones porcentuales - Datos como medidas de tendencia central, dispersión y posición - Selección de gráficas básicas adecuadas para la representación visual de datos	10 horas
4	Introducción al Modelado de Datos	- Conceptos básicos de modelado de datos - Importancia de modelado de datos dentro del proyecto de data lake y transformación digital de la PUCP	10 horas

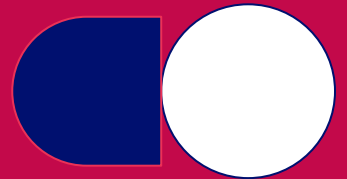
Tema	Sesión	Contenido
Monitoría Viernes 4:00 PM Zoom	1 (1 hr)	• Resolución de dudas de lo visto hasta el momento • Su asistencia no es obligatoria
	2 (1 hr)	• Resolución de dudas de lo visto hasta el momento • Su asistencia no es obligatoria
	3 (2 hr)	• 20 minutos por grupo para identificación del proyecto en el que van a trabajar
	4 (2 hr)	• 20 minutos por grupo para resolución de dudas y ayudar con procesamiento y/o visualización de datos
	5 (2 hr)	• 20 minutos por grupo para resolución de dudas y ayudar con procesamiento y/o visualización de datos
	6 (2 hr)	• 20 minutos por grupo para resolución de dudas y ayudar con procesamiento y/o visualización de datos

Evaluación	Descripción
Trabajo Final	Los estudiantes se organizan en grupos de 4 a 5 estudiantes. Seleccionan un tablero descriptivo existente y aplican mejoras utilizando los conceptos aprendidos en el curso. Es decir, implementan mejoras visuales y/o incluyen estadísticas descriptivas adicionales para enriquecer el dashboard. Este proyecto permite a los estudiantes aplicar habilidades recién adquiridas para optimizar la comprensión y la presentación visual de la información contenida en el tablero seleccionado. Presentación última sesión: 20 de agosto Máximo 6 grupos



SESIÓN I:

Introducción a Data Literacy





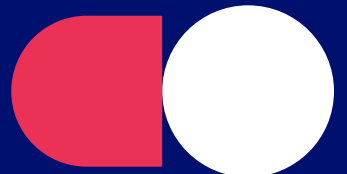
Al finalizar la sesión, lograrás:

- Definir data literacy y su impacto en el ámbito organizacional
- Describir la estructura de la pirámide DIKW y su utilidad en la comprensión de datos.



¿Qué es Data Literacy?

¿Qué creen que es?





Data Literacy



Nuestra capacidad para

leer,
trabajar,
analizar y
comunicarse con los datos

independientemente del rol de la persona y el nivel de
habilidad o herramientas que utilice

Data Literacy: otras definiciones



- The ability to recognize, evaluate, work with, communicate, and apply data in the context of business priorities and outcomes. Data literacy is the ability to read, work with, analyze, and argue with data

Forrester

- Ability to read, write and communicate data in context, including an understanding of data sources and constructs, analytical methods and techniques applied — and the ability to describe the use case, application and resulting value.

Gartner





Tanto los líderes empresariales como los empleados predicen que la alfabetización de datos será la habilidad más demandada para el año 2030.

QLIK (2023)

1,209 C-level executives and 6,197 global full-time employees in organizations of 50+ employees in the UK, USA, Germany, France, Japan, Australia and New Zealand in October and November 2021.



El 58 % de los empleados encuestados cree que la alfabetización de datos les ayudará a mantenerse relevantes en su rol ante el creciente uso de la inteligencia artificial.

Qlik (2023)



de cada 100 responsables de la toma de decisiones organizacionales, desde los directivos junior hasta la alta dirección, se sienten plenamente seguros de su capacidad para leer, trabajar, analizar y discutir con esos datos.

Qlik (2018,2023)



Se ha encontrado que las organizaciones con altos niveles de alfabetización en datos parecen generar entre un **tres y un cinco por ciento** más de capitalización de mercado, o un equivalente en dólares de entre 320 y 534 millones de dólares en valor empresarial, que aquellas que no los tienen.

The Data Literacy Project, "The Data Literacy Index," 2018.



¿Qué NO es Data Literacy?



- No es la capacidad de usar una herramienta o un conjunto de herramientas (Power BI, Excel)
- No es tener un conocimiento avanzado en estadística, matemáticas o programación
- No es convertirse en un científico de datos
- No es un fin en sí mismo
- No es solo para áreas técnicas



¿Cómo percibe que tener un nivel acorde de data literacy podría contribuir al logro de los objetivos de su área?

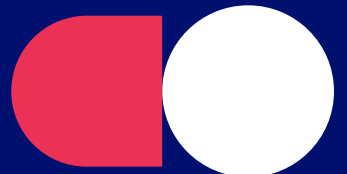


- slideswith.com/PSESIONSI_I





¿Qué involucra?





Conocimiento

Asimilación



Interpretación

**Escepticismo
constructivo**



1. Conocimiento

Qué son los datos, de dónde vienen, cómo se clasifican y por qué son importantes y entender los conceptos básicos que permiten interpretar resultados.



Característico de la organización

Cada organización tiene algunos términos que son únicos en relación con sus datos, indicadores y a la forma en cómo se organiza la información

Matemáticas y estadística

Conocimiento básico de conceptos como correlación, medidas de dispersión y de centralidad, diferencia entre un dato cualitativo y uno cuantitativo

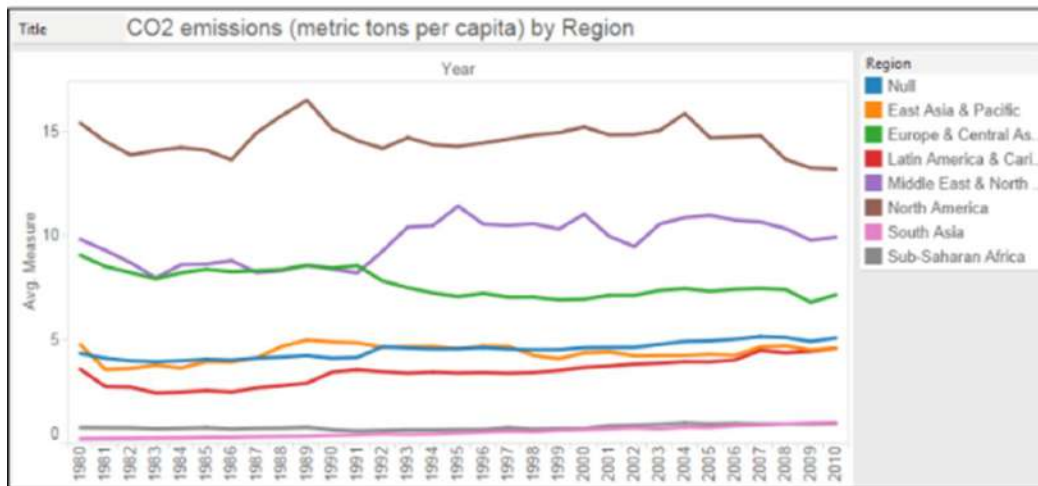


¿Tiene algún ejemplo?



2. Asimilación

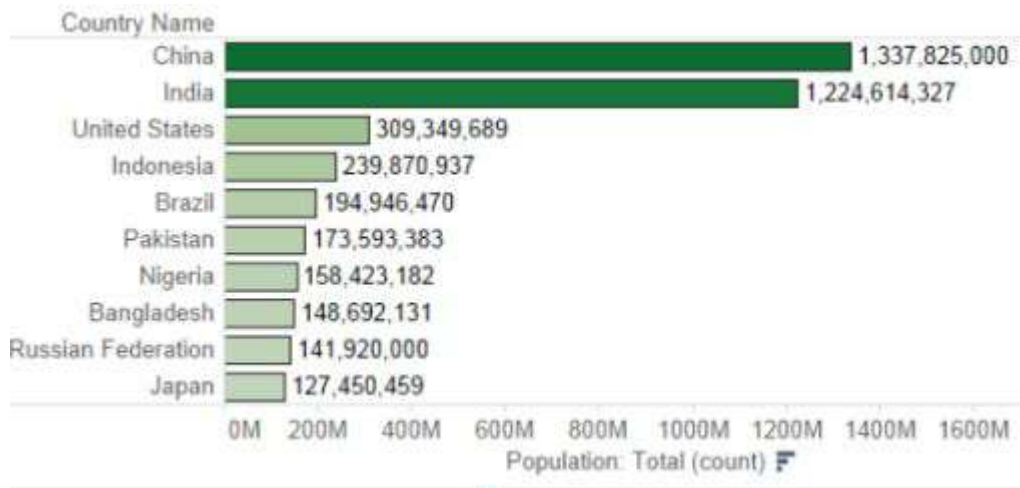
Es la capacidad de **recibir, procesar y contextualizar** la información proveniente de datos.



¿Tienen las tablas y gráficos
títulos claros y descriptivos?



2. Asimilación



¿Cuál es el periodo de análisis de los datos?



2. Asimilación



Country	1950	1951	1952	1953
Afghanistan	8151	8277	8407	8543
Africa	229895	234594	239501	244621
Albania	1215	1240	1269	1303
Algeria	8753	8953	9141	9326
American Sar	19	19	20	20
Andorra	6	7	7	8
Anguilla	4148	4219	4297	4377
Antigua and E	5	5	5	6
Argentina	17150	17506	17865	18224

¿Cuál es la fuente de la información?



2. Asimilación



¿Cuáles son escalas de los gráficos y son las apropiadas?



2. Asimilación

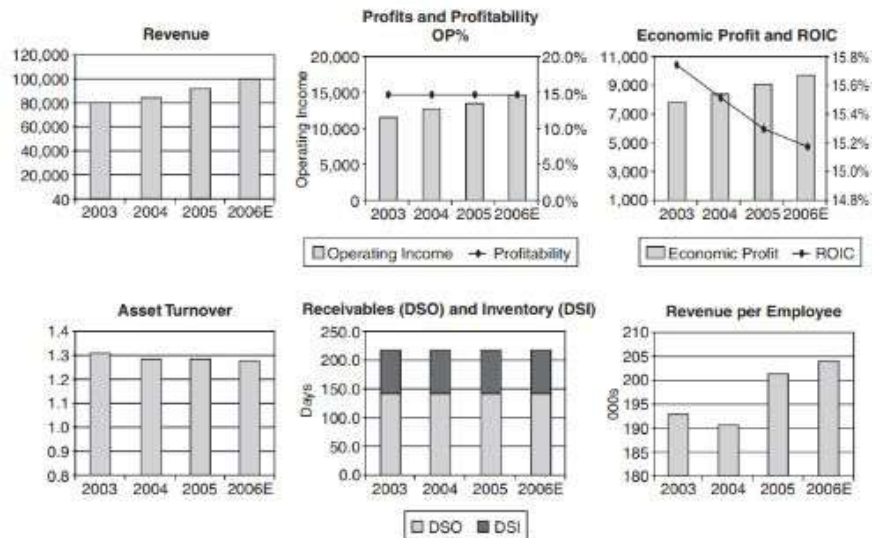
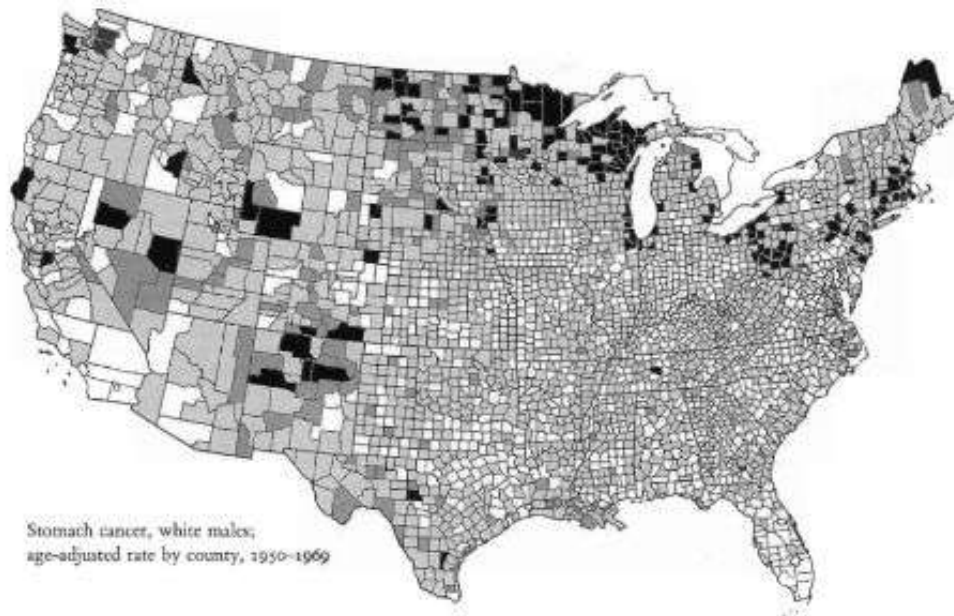


FIGURE 2.3 Key Performance Trends for Simple Co.

¿Entiendo las fórmulas detrás de las variaciones porcentuales, razones, entre otros?



2. Asimilación



¿Es claro cuál fue el criterio
para filtrar los datos?



3. Interpretación

Nuestra capacidad para darle sentido a los datos



¿Puedo identificar tendencias?



3. Interpretación



¿Puedo identificar
tendencias?

¿Puedo identificar patrones?



3. Interpretación



¿Puedo identificar
tendencias?

¿Puedo identificar patrones?

¿Tengo outliers?



3. Interpretación



¿Puedo identificar tendencias?	¿Puedo identificar patrones?	¿Tengo outliers?
¿Hay agrupaciones?		



3. Interpretación



¿Puedo identificar tendencias?	¿Puedo identificar patrones?	¿Tengo outliers?
¿Hay agrupaciones?	¿Hay valores perdidos?	



3. Interpretación



¿Puedo identificar tendencias?	¿Puedo identificar patrones?	¿Tengo outliers?
¿Hay agrupaciones?	¿Hay valores perdidos?	¿Puedo responder la pregunta de la organización con los datos?



4. Escepticismo y curiosidad

Nuestra capacidad para analizar con pensamiento crítico.

Sesgos

Image recognition systems that use biased machine learning data sets will inadvertently magnify that bias. Researchers are examining ways to reduce the effects.



COOKING	
ROLE	VALUE
AGENT	▶ WOMAN
FOOD	▶ PASTA
HEAT	▶ STOVE
TOOL	▶ SPATULA
PLACE	▶ KITCHEN



COOKING	
ROLE	VALUE
AGENT	▶ WOMAN
FOOD	▶ FRUIT
HEAT	▶ —
TOOL	▶ KNIFE
PLACE	▶ KITCHEN



COOKING	
ROLE	VALUE
AGENT	▶ WOMAN
FOOD	▶ MEAT
HEAT	▶ GRILL
TOOL	▶ TONGS
PLACE	▶ OUTSIDE



COOKING	
ROLE	VALUE
AGENT	▶ WOMAN
FOOD	▶ VEGETABLES
HEAT	▶ STOVE
TOOL	▶ TONGS
PLACE	▶ KITCHEN



COOKING	
ROLE	VALUE
AGENT	▶ MAN
FOOD	▶ —
HEAT	▶ STOVE
TOOL	▶ SPATULA
PLACE	▶ KITCHEN

In this example of gender bias, adapted from a report published by researchers from the University of Virginia and the University of Washington, a visual semantic role labeling system has learned to identify a person cooking as female, even when the image is male.



4. Escepticismo y curiosidad

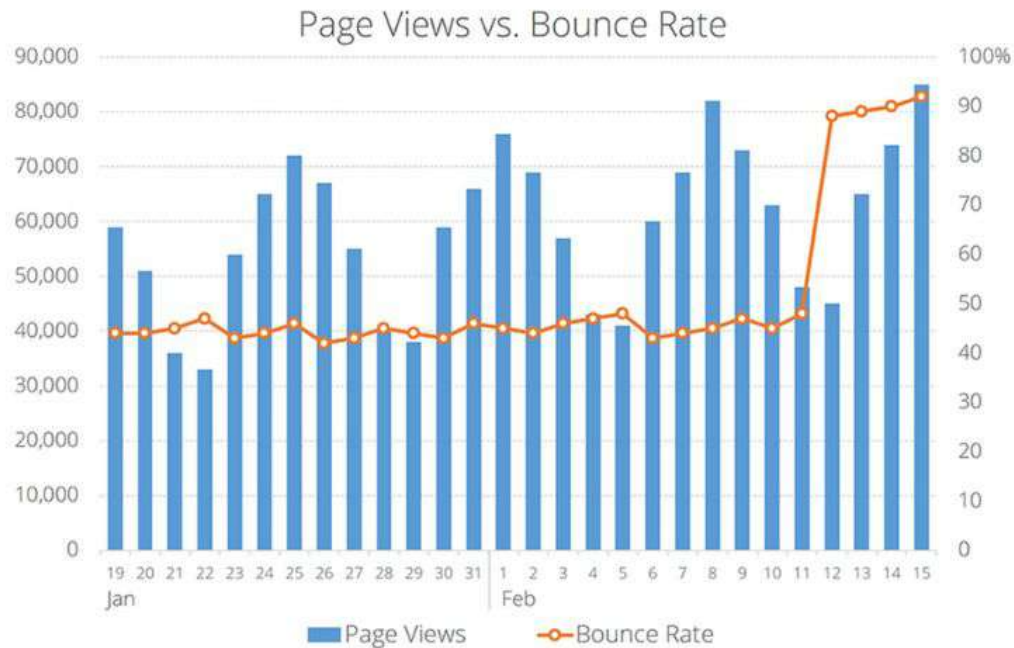


Supuestos

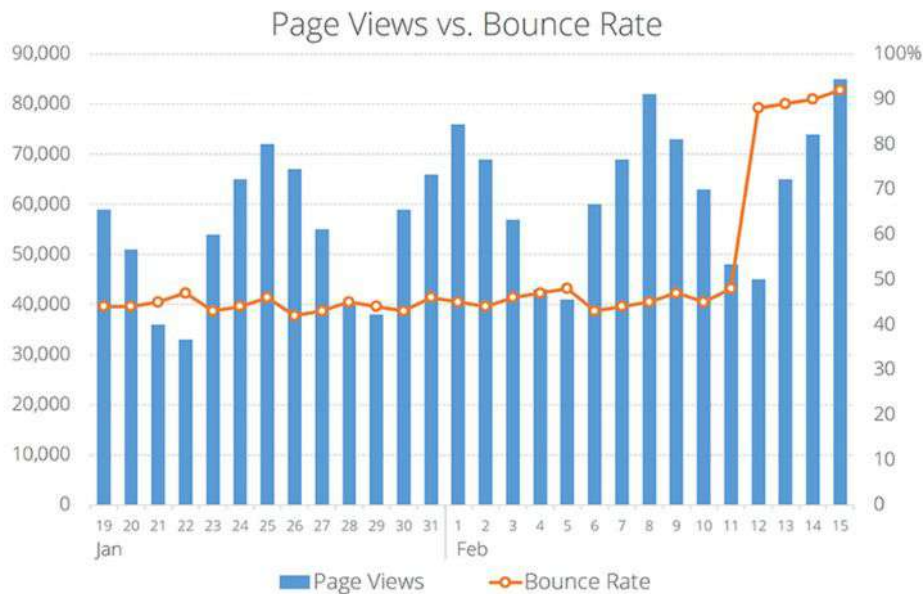
- COVID-19 has disrupted our behaviours, playing havoc with AI systems.
- Subject matter experts, who understand the context in which an AI system is operating, should assess how the COVID-19 crisis may have affected it.

MIT Technology Review [reported last week](#) that a number of AI systems are breaking down in the context of the COVID-19 pandemic. Systems designed for tasks ranging from inventory management to recommending online content are no longer working as they should.

Suponga que usted es un analista de marketing digital y le entregan esta visualización



Fuente:
<https://www.forbes.com/sites/brentdykes/2017/03/09/why-companies-must-close-the-data-literacy-divide/#4c7ae3ff369d>



Fuente:
<https://www.forbes.com/sites/brentdykes/2017/03/09/why-companies-must-close-the-data-literacy-divide/#4c7ae3ff369d>

¿Qué sería asimilación?

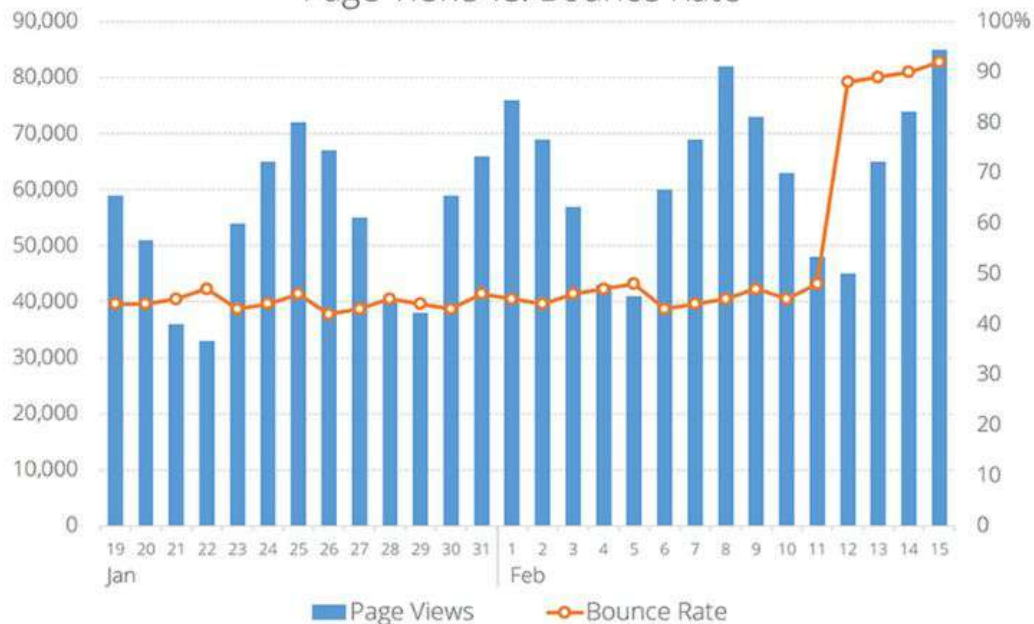
Entender la composición del gráfico los ejes, que hay dos indicadores en la misma gráfica, el periodo de análisis 28 días



Ejemplo



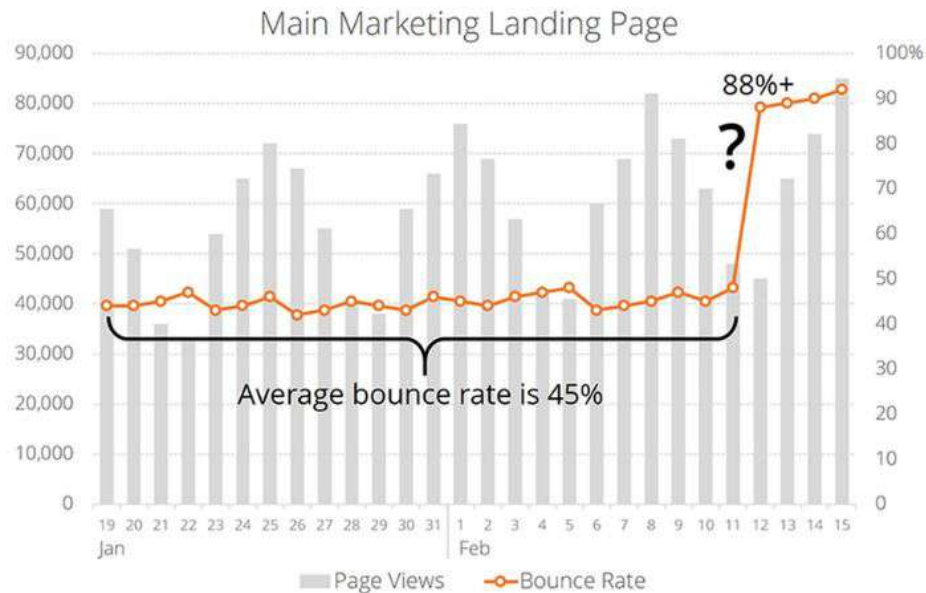
Page Views vs. Bounce Rate



Fuente:
<https://www.forbes.com/sites/brentdykes/2017/03/09/why-companies-must-close-the-data-literacy-divide/#4c7ae3ff369d>

¿Qué sería conocimiento?

Saber qué es Page Views y
qué es Bounce Rate, y
cómo se miden



¿Qué sería interpretación?



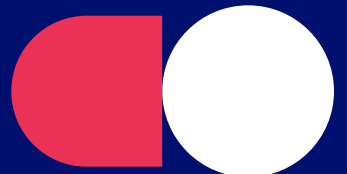
Actividad en Miro



- 1 Tamaño de grupos: de 6 personas
- 2 Descripción: en el gráfico asignado identifique qué sería conocimiento, asimilación, interpretación y curiosidad o escepticismo
- 3 Tiempo: 10 minutos



3. Pirámide DIKW



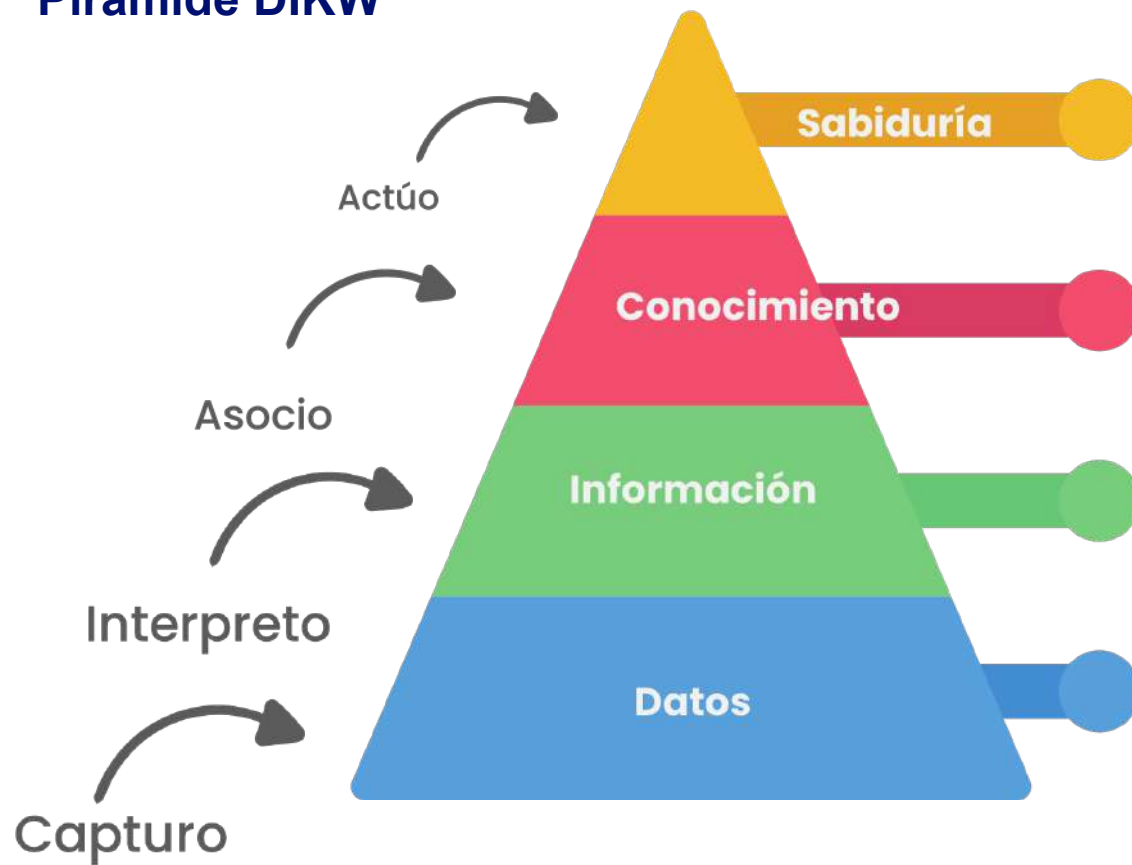


Pirámide DIKW





Pirámide DIKW





Pirámide DIKW





- Hechos, eventos, transacciones, etc., que han sido registrados
- Son el inicio o los cimientos de la comprensión -> Calidad



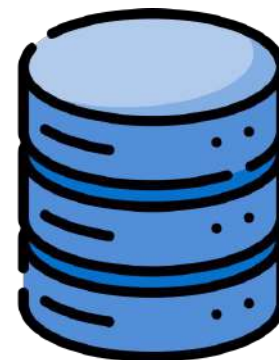
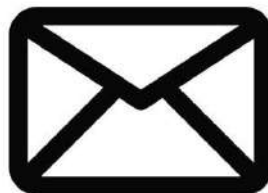
Ejemplos de dónde podemos encontrar datos



Datos



Pueden estar en un ambiente digital





Datos

Pueden estar en un ambiente físico



PUCP



FACTURA

LOGO

Reto Polo Paella Inc.
Carretera Muelle 36
37531 Avila, Avila

FACTURAR A
Leda Vilmaral
Virgen Blanca 63
08758 Burgos, Burgos

ENVIAR A
Leda Vilmaral
Corcon Bajos 68
47300 Cádiz, Cádiz

N° DE FACTURA EIS-001
FECHA 29/01/2019
N° DE PEDIDO 1730/2019
FECHA VENCIMIENTO 29/01/2019

CANT.	DESCRIPCIÓN	PRECIO UNITARIO	IMPORTE
1	Talla pequeña traje de luces en rojo	100.00	100.00
2	Mui grande chorrito	25.00	50.00
3	Equipo de Fútbol	5.00	15.00
Subtotal			165.00
IVA 21.0%			34.65
TOTAL			199.65 €

Laura Lora

Gracias

CONDICIONES Y FORMA DE PAGO

El pago se realizará en un plazo de 15 días

Banco Santander
IBAN: ES12 3456 7891
SWIFT/BIC: ABCDESMXXX



- Son datos organizados o procesados para que nos resulten legibles y útiles.
- Son datos dentro de un contexto
- Nos ayudan a resolver:
 - Qué
 - Quién
 - Dónde
 - Cuándo





- Resulta de analizar e interpretar la información para encontrar patrones, asociaciones, tendencias.
- Nos ayuda a responder preguntas como:
 - ¿Cómo? ¿Por qué?
- ¿A qué está asociada la información?
- ¿Qué comparaciones podemos hacer para comprender mejor la pertinencia de los datos?





- Aplicación de los conocimientos que nos hemos formado a partir de la información recolectada de los datos.
- Es la toma de decisiones y acciones correctas o inteligentes. Se evita hacer ideas equivocadas con los datos y realizar las cosas objetivamente (a pesar de la complejidad y las contradicciones).





Ejemplo: Área contable



- **Datos:** transacciones de ventas, gastos, valoración de activos, pago de nómina
- **Información:** Estados financieros y métricas de ventas, flujo de cada, utilidad bruta, etc.
- **Conocimiento:** encontrar patrones de estacionalidad en las ventas y cómo esto puede afectar ingresos y gastos
- **Sabiduría:** Gerente Financiero toma decisiones informadas para priorizar inversiones, asignar presupuestos, entre otros para maximizar la rentabilidad de la compañía





- slideswith.com/PSESIONSI_2





Respuestas



- ¿Cuáles son los cuatro niveles de la pirámide DIKW? R/
Datos-Información-Conocimiento-Sabiduría
- Suponga que tiene unos objetivos de estado físico y está comparando número total de km que corrió la semana pasada con los km que corrió en semanas anteriores desde que empezó a hacer ejercicio. ¿Este es un ejemplo de que nivel de la pirámide?
R/ Convertir la información en conocimiento mediante la asociación
- Es posible tener información sin datos, pero no es posible tener datos sin información. R/ Falso
- Si te encuentras con una cifra como, por ejemplo, 24.901, pero no se indican las unidades, ¿en qué nivel de la pirámide DIKW se quedará esta cifra hasta que determines las unidades y lo que significan? R/ Datos
- Cuánto más ascendemos en la pirámide DIKW, más "elemento humano" (suposiciones, interpretaciones, asociaciones, prioridades y creencias) añadimos a la situación. R/ Verdadero



Próxima sesión



- Trae un archivo de Excel o Google Sheets con al menos 10 filas de datos que uses frecuentemente en tu trabajo. Puede ser una base de clientes, asistencia, productos, encuestas, indicadores, etc.
- Si no puedes traer un archivo real por confidencialidad, crea una tabla simulada con datos representativos.





Gracias

