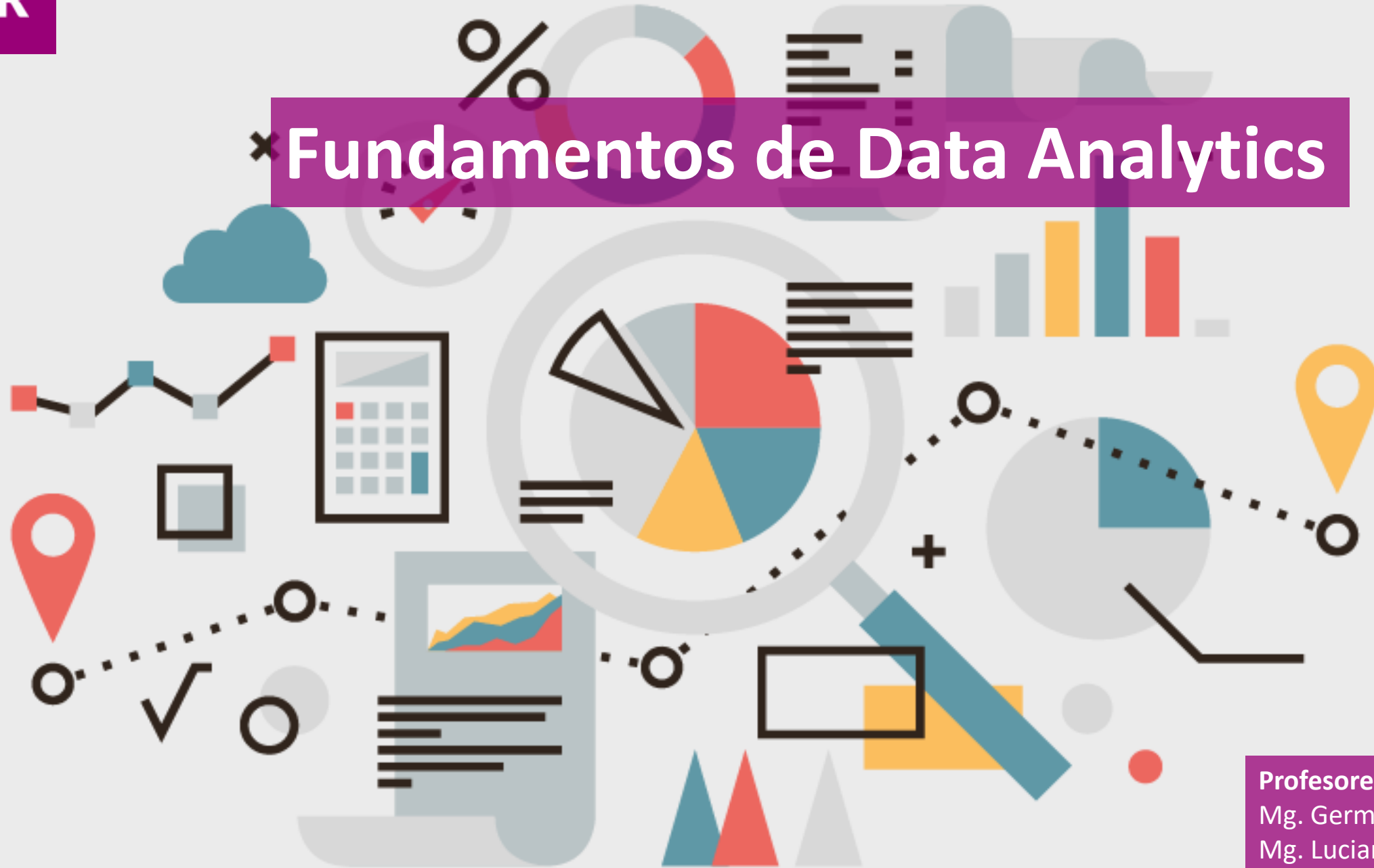


× Fundamentos de Data Analytics



Profesores

Mg. Germán Tessmer

Mg. Luciano Jara Musuruana

Fundamentos de Data Analytics

Programa

- Fundamentos de estadística
- Gráficos y tablas para la descripción de datos Excel
- Medidas resumen en Excel
- Introducción al sistema de consulta de bases de datos SQL

¿Por qué Excel?

- Lingua franca
- Building blocks (si, vamos a estar “switchcando” bastante)
- Percepción de oferta laboral en Rosario



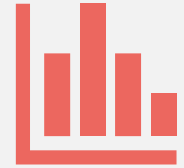
Antes de empezar... el experimento de Linda

Linda tiene 31 años, es soltera, no tiene pelos en la lengua y es muy inteligente. Se licenció en Filosofía. Cuando estudiaba, le preocupaban mucho los problemas de discriminación y justicia social, y también participó en manifestaciones antinucleares

¿A que categoría pertenece?

1. Linda trabaja de cajera en un banco
2. Linda es profesora de una escuela primaria
3. Linda es vendedora de seguros
4. Linda trabaja de cajera en un banco y es un miembro activo del movimiento feminista

Análisis de datos y toma de decisiones



- El análisis de datos es clave en la era digital.
- Permite convertir datos en información valiosa.
- Identifica patrones y tendencias ocultas.



- Toma de decisiones basada en evidencia y datos.
- Identificación de oportunidades y riesgos.
- Mejora de la eficiencia y la productividad.
- Personalización y segmentación.
- Innovación y ventaja competitiva.

Ejemplo

En el marco del Plan Ganadero 2030, se estipula crear y fortalecer siete centros de recría. Hay cuatro de esos centros que ya están en funcionamiento: el Establecimiento Ñanilda/Cre-Nor (Margarita Belén), El Timbozal (Puerto Vilelas), Don Ramón (Presidencia Roque Sáenz Peña) y Las Dos T (Cote Lai). “Estos programas se ponen en marcha de acuerdo a las necesidades del sector. Esto implica llegar al pequeño productor con el apoyo financiero y la posibilidad de generar la recría de sus terneros”, señaló el Ministro de Producción. “Se estipula que tendrán una capacidad de trabajo de 3.800 cabezas de ganado en total”.

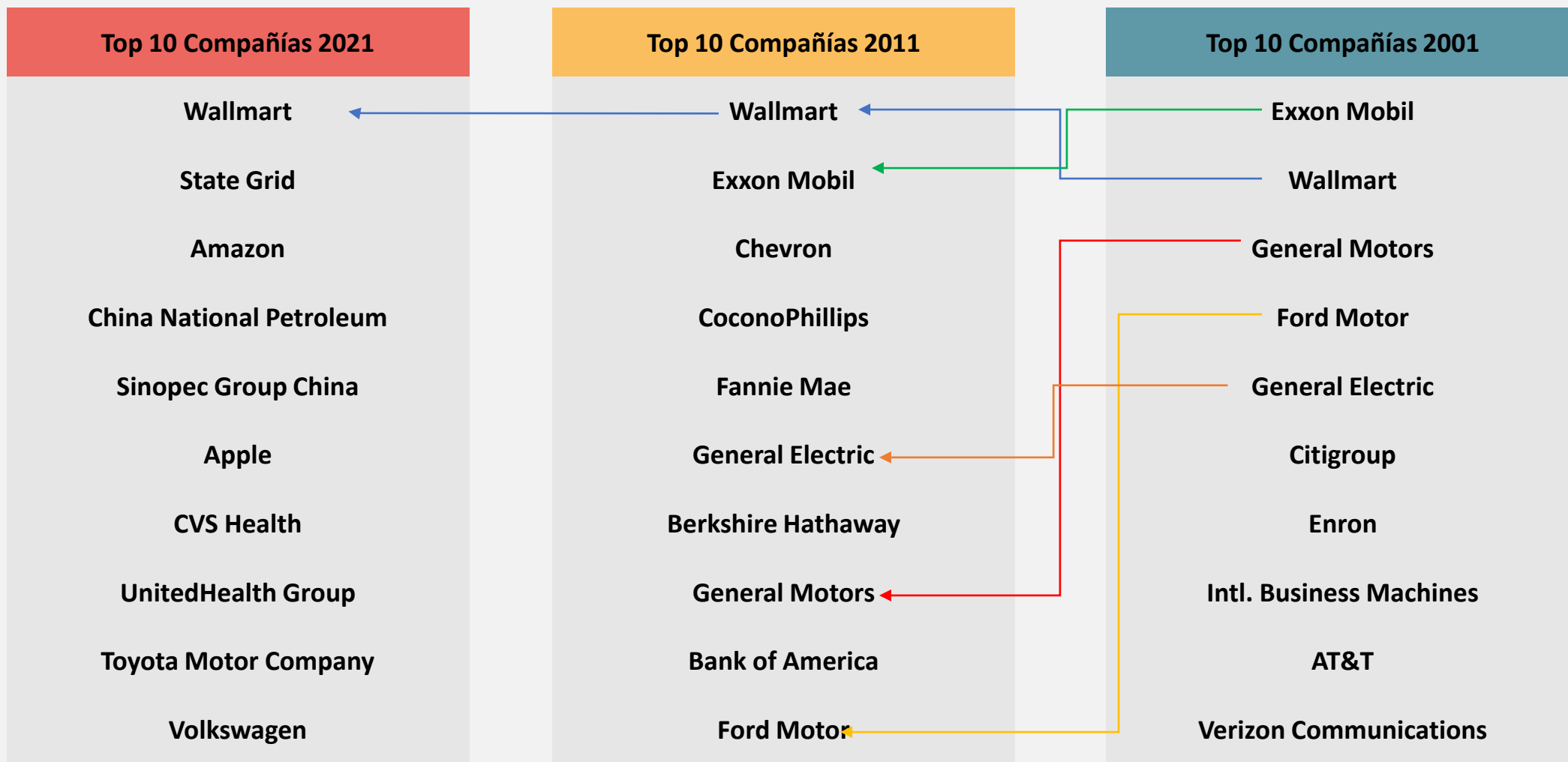
Departamento	Establecimientos		Unid. Produc.		Total Bovinos	
	Cant.	%	Cant.	%	Cant.	%
Cmte Fernandez	1	5,56%	2	4,35%	32	0,23%
San Fernando	1	5,56%	1	2,17%	201	1,46%
Primero de Mayo	1	5,56%	1	2,17%	463	3,37%
Total	18	100,00%	46	100,00%	13.759	100,00%



La **transformación digital** es el fomento de evolución y nuevos modelos de negocio incorporando la digitalización de archivos e incorporando lo digital a todas las áreas de negocio.

– Harvard Business Review, 2021

Transformación digital, un ejemplo



Organización Data Driven (I)

- Es una disciplina que utiliza diversas técnicas y herramientas de análisis para aprovechar los datos generados dentro de un ámbito o empresa para su beneficio y el de sus clientes.
- Dicho de otro modo, se trata de sacar valor a los millones de datos de los que hoy disponemos para tomar mejores decisiones basadas en ellos.



Datos

Los clientes compran el producto de datos una vez y continúan usándolo tal como está.



Refinamiento

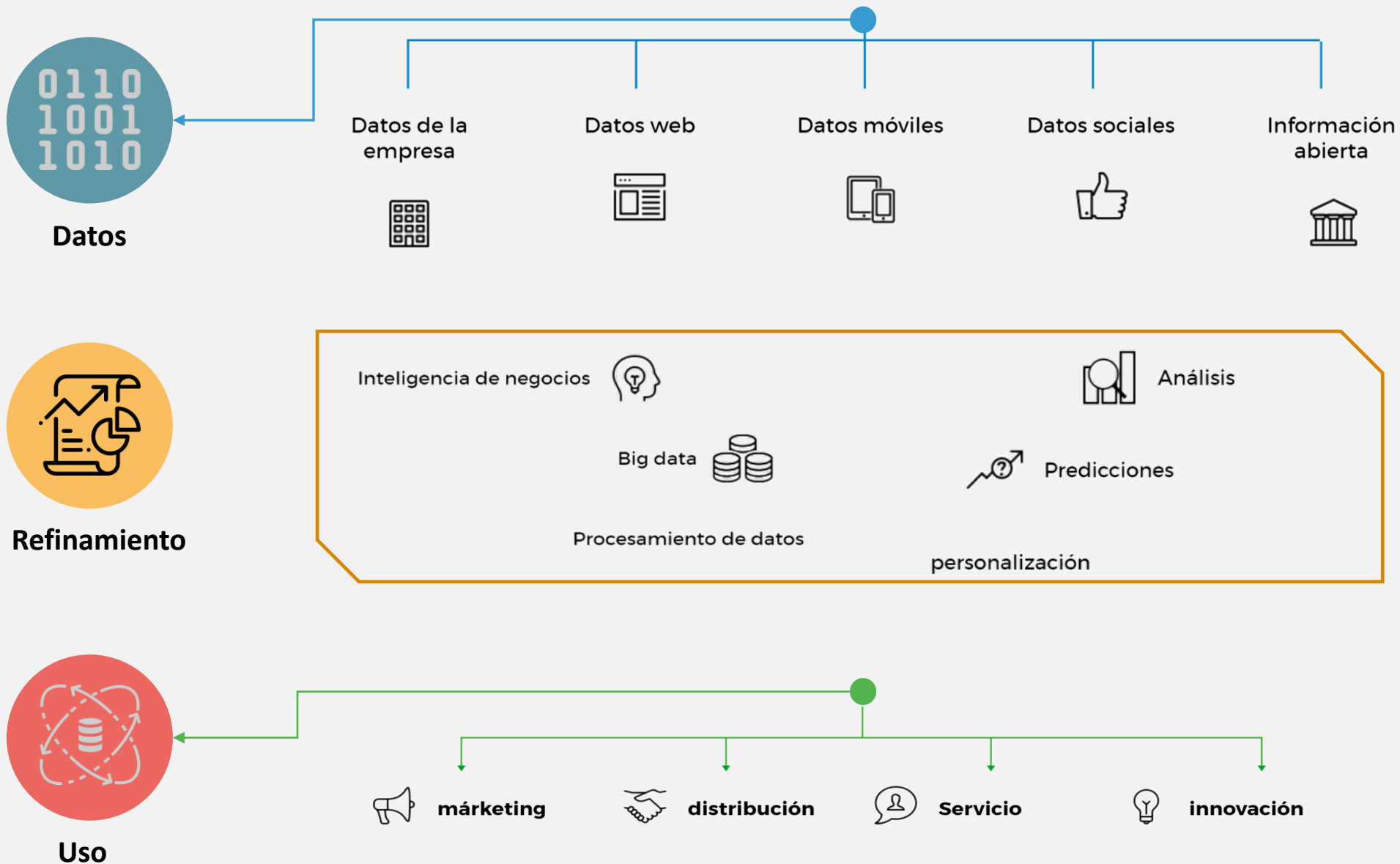
Un producto de datos es una aplicación informática que toma entradas de datos y genera salidas, devolviéndole al entorno.



Uso

Un producto de datos es información digital que se puede comprar y usar.

Organización Data Driven (I)



¿Qué es la estadística? Una introducción

- La **estadística** es una disciplina que se centra en la recopilación, análisis, interpretación y presentación de datos numéricos.
 - Utilizamos técnicas estadísticas para realizar inferencias, identificar patrones, evaluar relaciones y medir la confiabilidad de nuestros resultados.
 - La estadística proporciona un marco teórico sólido para el diseño de experimentos, la elaboración de modelos y la validación de hipótesis en el análisis de datos.
 - Además, nos ayuda a comprender la probabilidad y las distribuciones de datos, lo cual es esencial para realizar predicciones y estimaciones.
-
- Es fundamental en el campo del **data analysis** debido a su capacidad para ayudarnos a comprender y tomar decisiones basadas en los datos.
 - Es decir, el proceso de analizar datos para extraer conocimientos y tomar decisiones informadas.

Enfoque clásico y enfoque bayesiano

Enfoque clásico:

- Se basa en la teoría de la probabilidad frecuentista.
- Considera que las probabilidades son objetivas y se pueden estimar a partir de la frecuencia relativa de eventos en muestras grandes.
- Utiliza métodos como la estimación de máxima verosimilitud y los intervalos de confianza.
- Se enfoca en estimar parámetros poblacionales y realizar inferencias sobre ellos.
- No tiene en cuenta información previa o subjetiva.

Enfoque bayesiano:

- Considera que las probabilidades son subjetivas y se pueden actualizar a medida que se obtiene nueva información.
- Utiliza el teorema de Bayes para combinar la información previa (distribución a priori) y la información de los datos (verosimilitud) y obtener la distribución a posteriori.
- Se enfoca en estimar la incertidumbre y la probabilidad de diferentes valores de los parámetros.
- Permite incorporar conocimientos previos y actualizarlos con evidencia empírica.

Enfoque clásico y enfoque bayesiano

Diferencias y consideraciones:

- El enfoque clásico se centra en la frecuencia relativa de los eventos y se basa en muestras grandes, mientras que el enfoque bayesiano se basa en la probabilidad subjetiva y permite actualizar las creencias con nueva información.
- El enfoque clásico proporciona estimaciones puntuales y intervalos de confianza, mientras que el enfoque bayesiano proporciona distribuciones de probabilidad a posteriori.
- El enfoque clásico no considera información previa, mientras que el enfoque bayesiano permite incorporar conocimientos previos.
- La elección entre enfoque clásico y enfoque bayesiano depende del contexto, las creencias subjetivas y los objetivos del análisis estadístico.

Big data, ¿qué es?

Small data

- Pocos datos
- Estructurados
- De generación lenta
- Caros

Big data

- Muchos datos
- No estructurados
- De generación inmediata
- Relativamente baratos



Funes (*el memorioso*) es big data sin estadística

– Stephen Stigler, 2016

Data mining, ¿qué es?

Es la extracción de información para descubrir patrones que sean válidos, novedosos, potencialmente útiles y comprensibles.

- Válidos: Tienden a mantenerse
- Novedosos: No lo conocíamos de antemano
- Útiles: Podemos idear acciones a partir de ello
- Comprensible: Podemos interpretar y comprender los patrones



Machine learning, ¿qué es?

Es el descubrimiento de esos patrones, que denominaremos modelos, el cual aprende de manera dinámica con el objetivo de generar la mejora predicción posible

