



UNIVERSIDAD
DE SANTIAGO
DE CHILE

DIPLOMADO EN

➤ **PRODUCT MANAGER**

ÁREA DE MARKETING



MÓDULO II:

DATA SCIENCE PARA
EL PRODUCT
MANAGER

**FACULTAD DE
ADMINISTRACIÓN Y ECONOMÍA**
DEPARTAMENTO DE ADMINISTRACIÓN

¡Bienvenidos!

Presentación

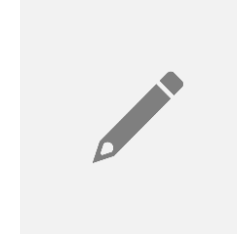
- Educación: Ingeniero Civil Industrial UC
- Experiencia: +7 años liderando equipos de Data Science, Marketing Analytics y Business Intelligence
- Capacitaciones adicionales: Bootcamp Python + Certificado en Inteligencia Artificial y aplicaciones en la empresa de UC Berkeley
- Herramientas: Python (PL), Power BI (BIS), GCP (CCP) y G-Suite (OS)
- Módulos: Data Science, Modelos Analíticos y Business Intelligence
- Rubros: Banca y Consumo Masivo





Agenda

Sesión #1



Introducción al *Data Science*

1. ¿Qué es el Data Science?
2. Proceso de Data Science
3. El Equipo de Data Science
4. Origen de los datos
5. Zoom: Bases relacionales y datos estructurados
6. Flujo de datos
7. Barreras de adquisición y tratamiento de datos
8. Ejercicio

Sobre el curso



- Qué es el Data Science
- Cómo opera en las distintas compañías
- Qué problemáticas se buscan resolver de la mano del Data Science
- Cómo se abordan dichas problemáticas – Modelos más utilizados
- Principales herramientas que ayudan al negocio
- Las “*buzzwords*” de la especialidad

Evaluaciones

- Evaluación #1: Abordar problemática típica de marketing a través de interpretación de datos y modelos fundamentales del Data Science
- Evaluación #2: Presentación de metas en base a modelos predictivos



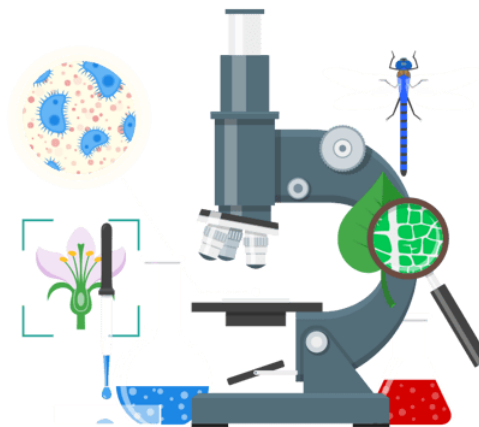
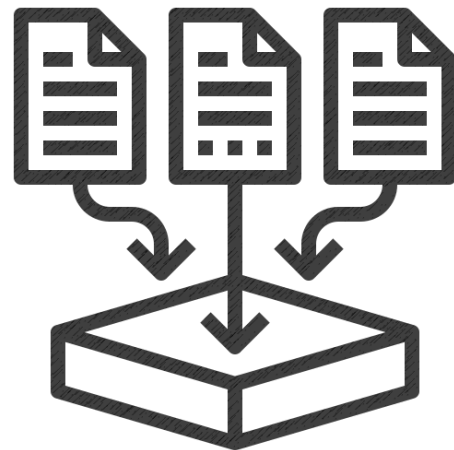
¿Qué es el Data Science?

DIPLOMADO EN

 **PRODUCT
MANAGER**

ÁREA DE MARKETING

Data Science



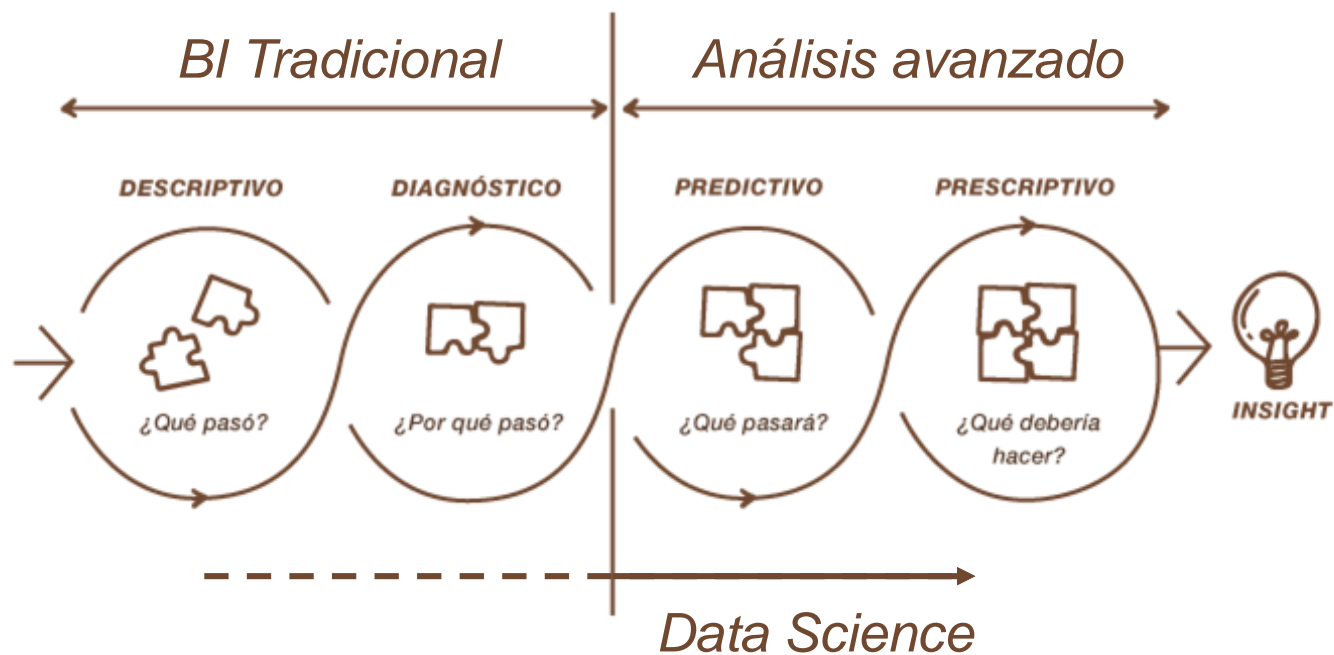
Un poco de contexto...

DIPLOMADO EN

**PRODUCT
MANAGER**

ÁREA DE MARKETING

EVOLUCIÓN DEL ANÁLISIS



Importancia de los Datos

Harvard
Business
Review

Technology And Analytics | How Data-Driven Decisions Help Restaurants Stay

DIPLOMADO EN

 **PRODUCT
MANAGER**


ÁREA DE MARKETING

Forbes

Member Full

FORBES > INNOVATION

Why High-Quality And Relevant Data Is Essential In Today's Business Landscape




Jeff White

Forbes Councils Member

Forbes Technology Council

COUNCIL POST | Membership (Fee-Based)



Apr 17, 2023, 08:45am EDT

Founder and CEO of [Gravy Analytics](#), a leading provider of enterprise location intelligence.

How Data-Driven Decisions Help Restaurants Stay Competitive

Debjit Roy, Eirini Spiliotopoulou, and Jelle de Vries

October 03, 2023



...the essential and key predictions of this multidisciplinary field

Updated: Nov 24, 2023, 04:56:00 PM IST

IES | Careers

St

New

IES ePaper

ids Tech

Careers

Opinion

NRI

Panache

ET TV


5


disciplinary field


ie essential and key

predictions of this multidisciplinary field

Updated: Nov 24, 2023, 04:56:00 PM IST









FOLLOW US

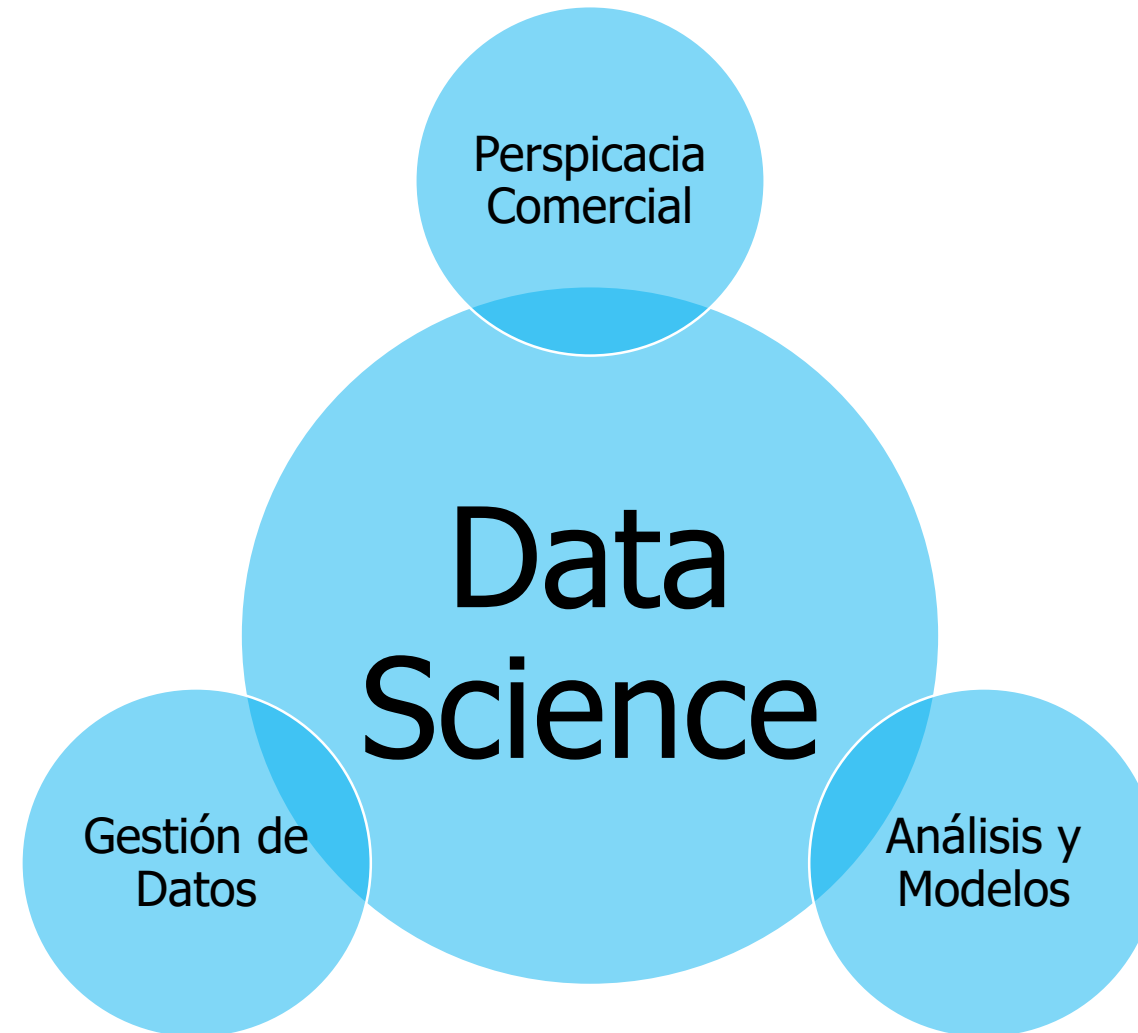
SHARE

FONT SIZE

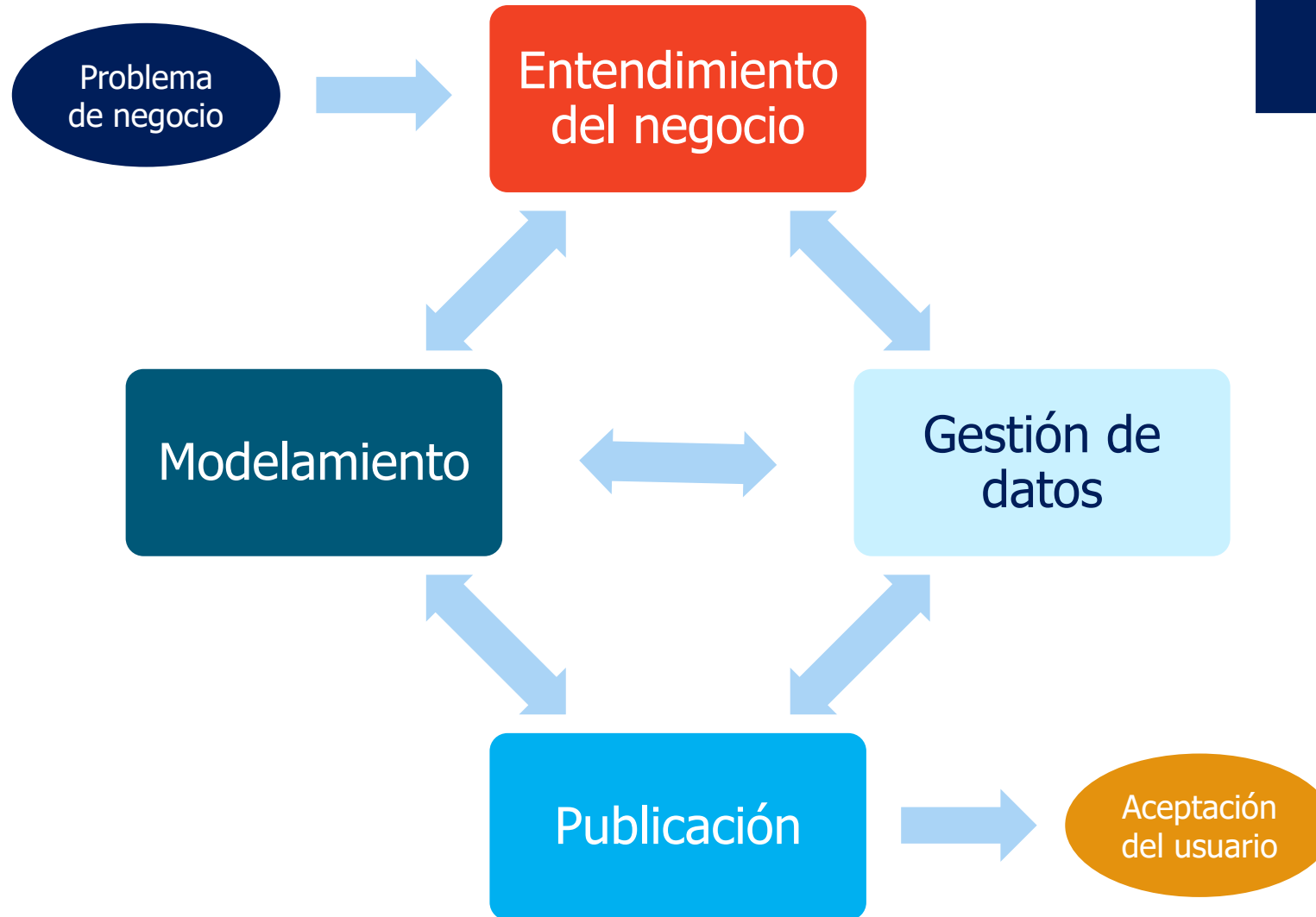




¿Qué es el Data Science?



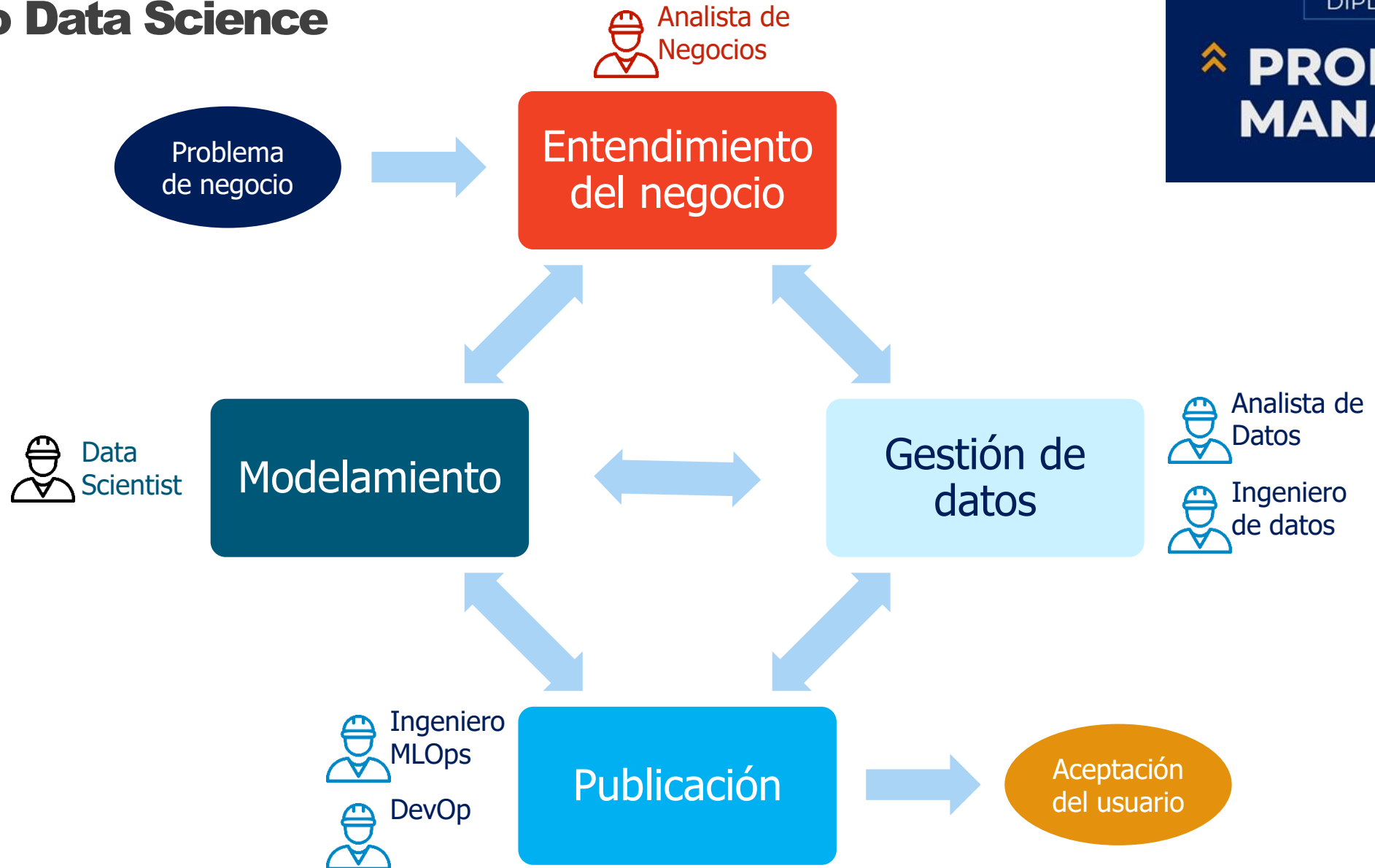
Proceso Data Science



Proceso Data Science Detallado

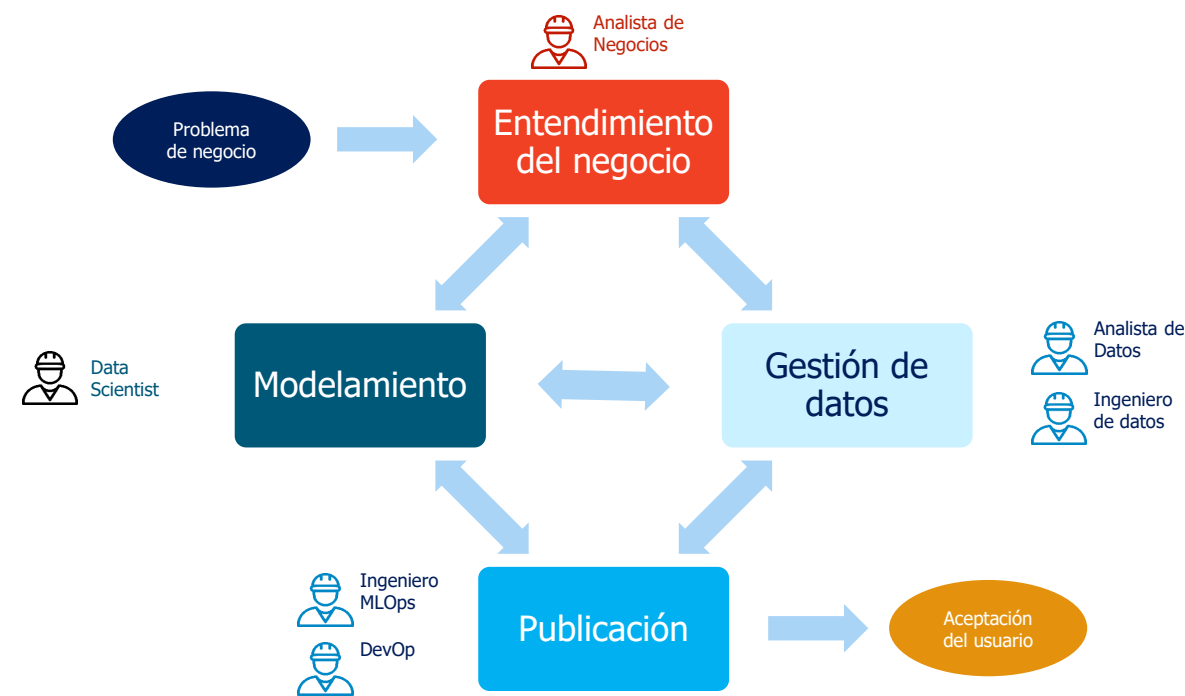


Proceso Data Science Roles



El equipo de Data Science

Habilidades



	Data Analyst	Machine Learning Engineer	Data Engineer	Data Scientist
Programming Tools	Very important	Very important	Very important	Very important
Data Visualization and Communication	Very important	Somewhat important	Somewhat important	Very important
Data Intuition	Somewhat important	Very important	Somewhat important	Very important
Statistics	Somewhat important	Very important	Somewhat important	Very important
Data Wrangling	Not that important	Not that important	Very important	Very important
Machine Learning	Not that important	Very important	Not that important	Very important
Software Engineering	Not that important	Somewhat important	Very important	Somewhat important
Multivariable Calculus and Linear Algebra	Not that important	Very important	Not that important	Somewhat important

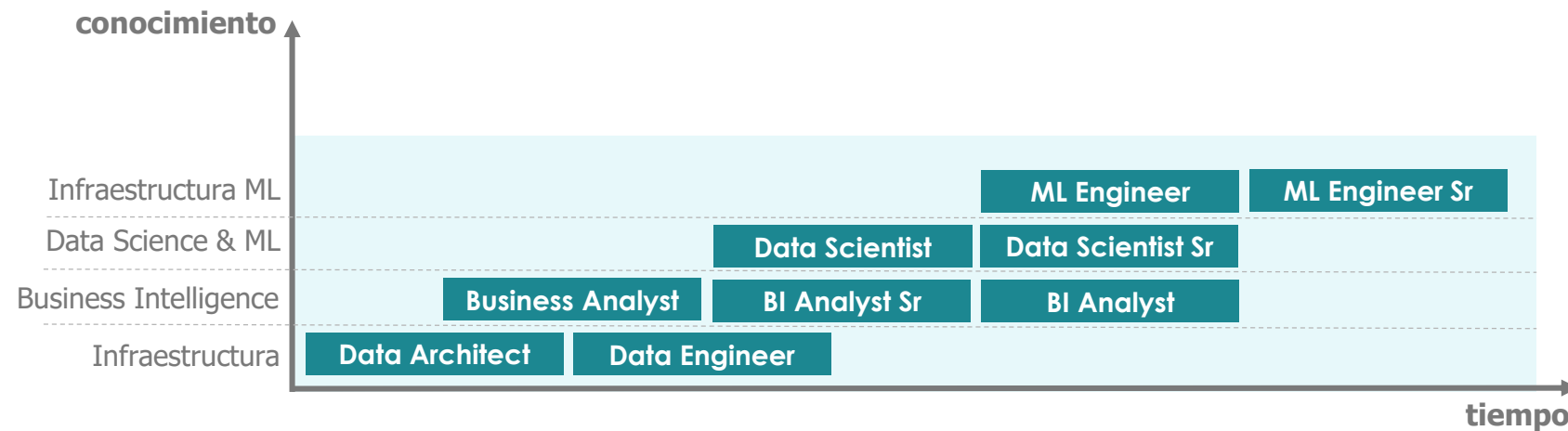
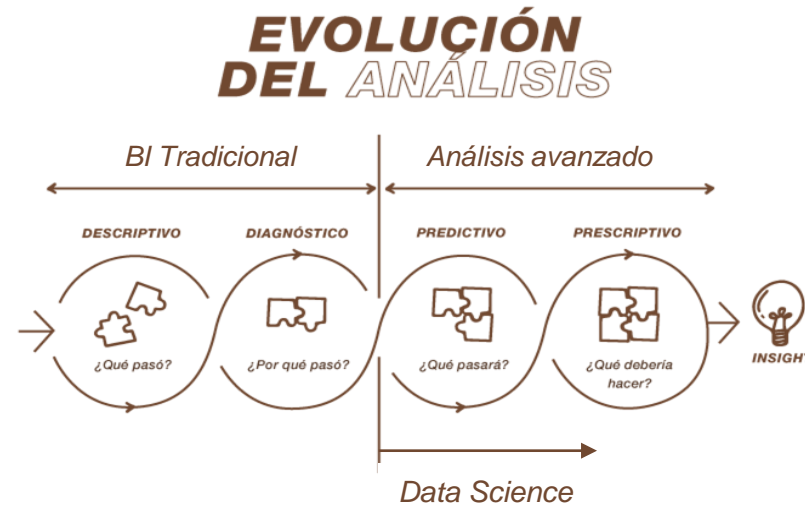
Not that important

Somewhat important

Very important

El equipo de Data Science

Funcionamiento en una empresa



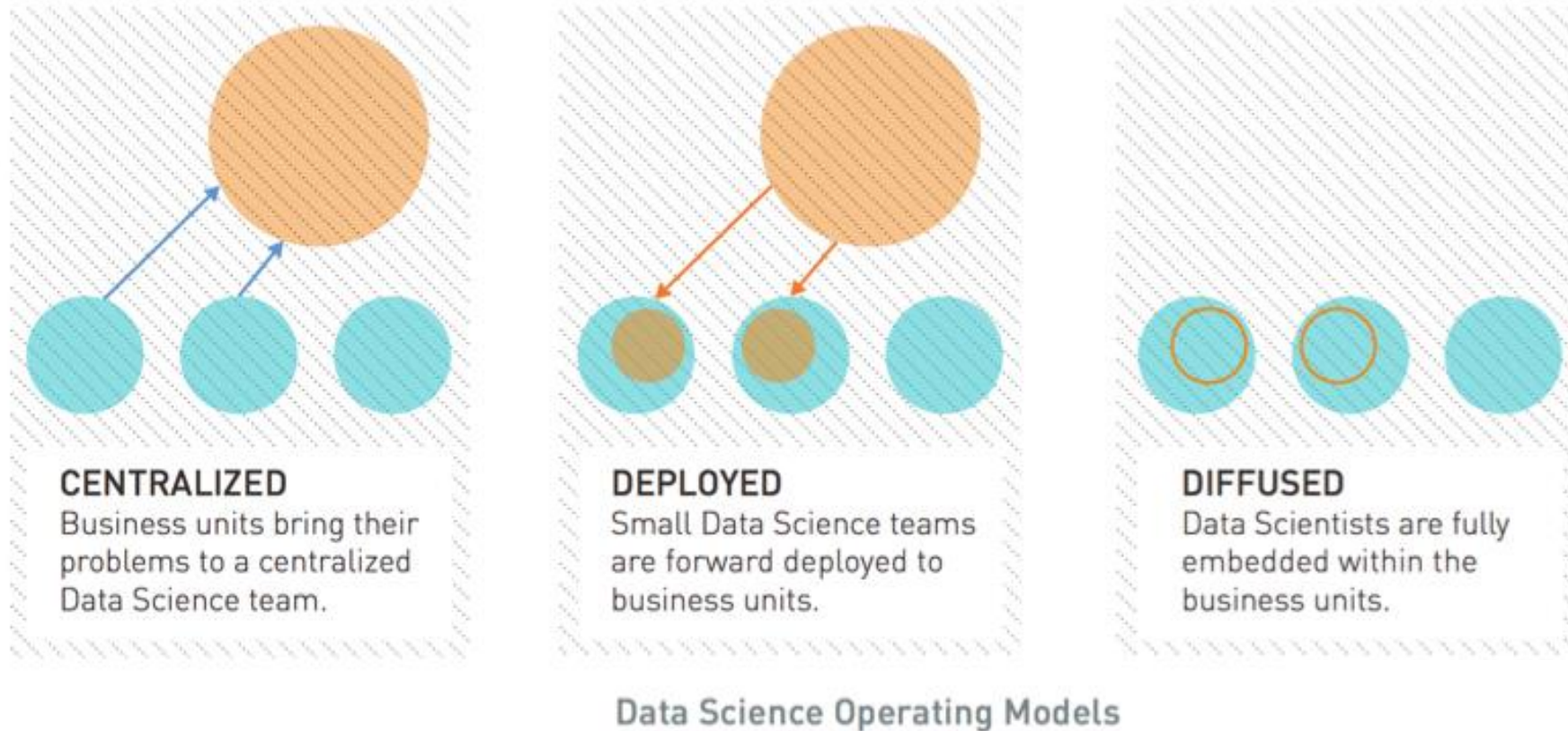
El equipo de Data Science

Funcionamiento en una empresa

DIPLOMADO EN

**PRODUCT
MANAGER**

ÁREA DE MARKETING



Datos como medio

Datos como objetivo

El equipo de Data Science

Funcionamiento en una empresa

DIPLOMADO EN

 **PRODUCT
MANAGER**

ÁREA DE MARKETING

Explore

Build

Run

EQUIPO DE EXPLORACIÓN



Squad
Lead



Analytics
Translator



Developers
(DS, DE, MLE, DV)

EQUIPO DE PRODUCTO



Squad
Lead



Analytics
Translator



Data Scientist



Roles Flotantes
(DE, MLE, DV)

ENABLERS



Technical
Lead



Agile
Lead



Translator
Lead



Squad
Lead



Analytics
Translator



Developers
(DS, DE, MLE, DV)



Data
Engineer



Staff Data
Scientist

En Chile...

DIPLOMADO EN

 **PRODUCT
MANAGER**

ÁREA DE MARKETING



Carlos Montes 30 SEP 2021 02:13 PM · Tiempo de lectura: 8 minutos

La ciencia de la transformación digital en la academia

Varias universidades en el país impactan en el área. Su aplicación y potencia artificial, ya es una realidad en muchos casos, ya sea por Covid-19 o el Trabajo Remoto.

Cambio Climático
Iniciativa de cambio climático
Fundación Data On
los efectos del calentamiento global implementados.

Por: Valentina Morales
04:00 hrs.

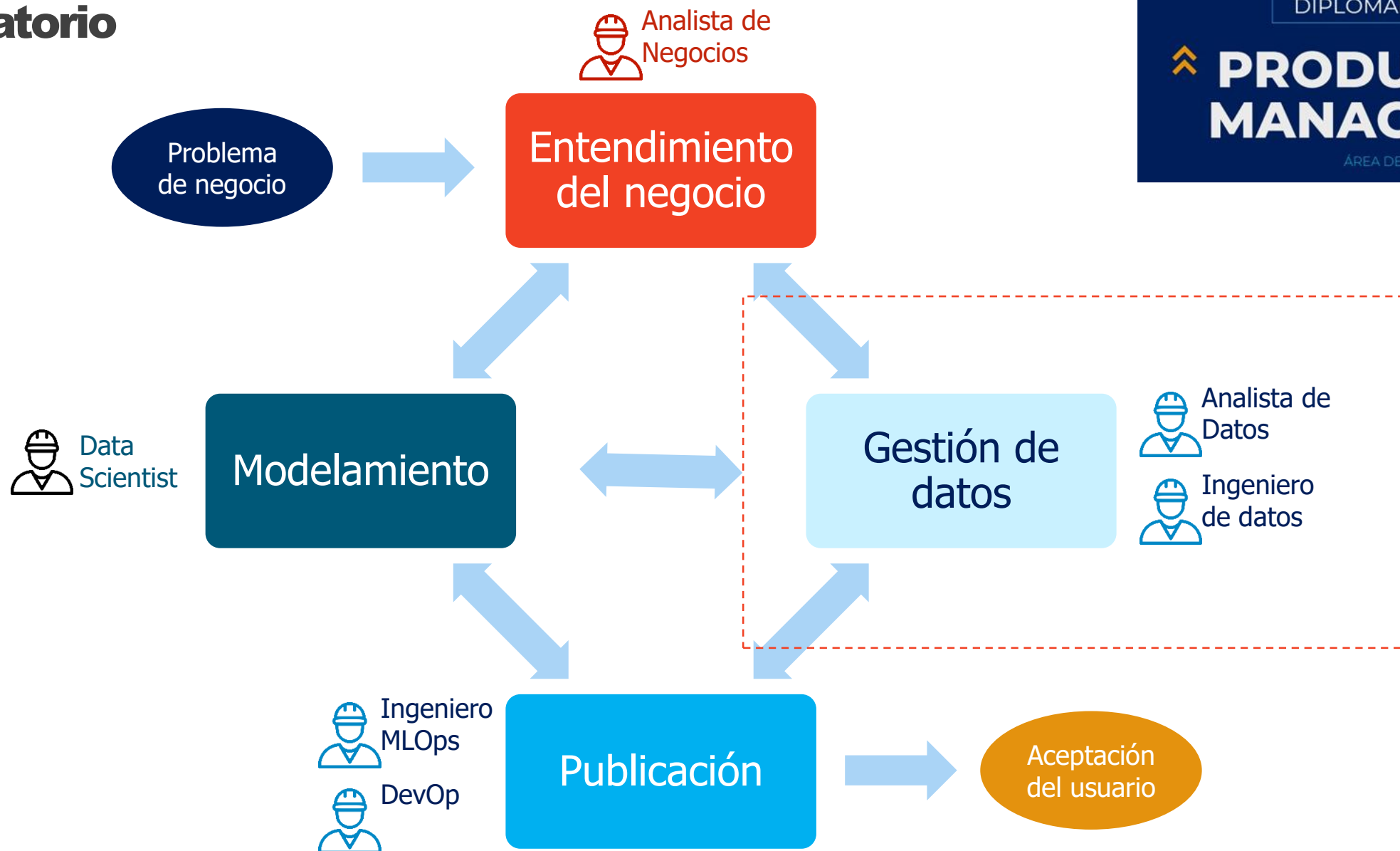
NOTICIAS

La relevancia de los Data Scientist para enfrentar los desafíos de la transformación digital



iBreak!

Recordatorio



Origen de los datos

Tipos

Bases de datos

- Suelen corresponder a data interna
- Suelen ser relacionales
- Contienen en su mayoría datos transaccionales (CRM o ERP)
- Requieren en general procesos ETL

Archivos planos

- Suelen corresponder a data externa
- Suelen tener formato XLS, CSV o JSON

Servicios web

- Suelen corresponder a data externa
- Se recurre en general a técnicas de web scraping (APIs, web crawler, ready-to-use crawler)

Otras fuentes

- Incluyen RSS feeds (Google Analytics)

Origen de los datos

Bases de datos

DIPLOMADO EN

 **PRODUCT
MANAGER**

ÁREA DE MARKETING

Bases de datos

- Suelen corresponder a data interna
- Suelen ser relacionales
- Contienen en su mayoría datos transaccionales (CRM o ERP)
- Requieren en general procesos ETL

Tipos de datos

ERP:

- Ventas
- Costos
- Márgenes
- Gastos

CRM:

- Datos generales de clientes actuales
- Necesidades, preferencias o gustos
- Datos sobre clientes potenciales
- Opiniones y comentarios de clientes

Consideraciones

Debido a que suelen ser de gran volumen, requieren una infraestructura y procesamiento avanzados, además de un tratamiento que apunte a su optimización, tanto de estructura como de consulta

Origen de los datos

Archivos planos

Archivos planos

- Suelen corresponder a data externa
- Suelen tener formato XLS, CSV o JSON

Tipos de datos

Competidores (Servicios de información):

- Ventas competencia
- Precios competencia
- Promociones competencia

Mercado:

- Estudios de mercado
- Encuestas

Consideraciones

- Dependiendo del volumen, pueden requerir conversión a bases de datos

y/o

- Dependiendo del carácter estratégico, pueden requerir integración a bases de datos internas

Origen de los datos

Servicios web

DIPLOMADO EN

 **PRODUCT
MANAGER**

ÁREA DE MARKETING

Servicios web

- Suelen corresponder a data externa
- Se recurre en general a técnicas de web scraping (APIs, web crawler, ready-to-use crawler)

Tipos de datos

Competidores (información web):

- Compras competencia
- Precios competencia
- Promociones competencia
- Presencia competencia

Mercado:

- Indicadores macroeconómicos
- Tipos de cambio
- Tasas de interés

Consideraciones

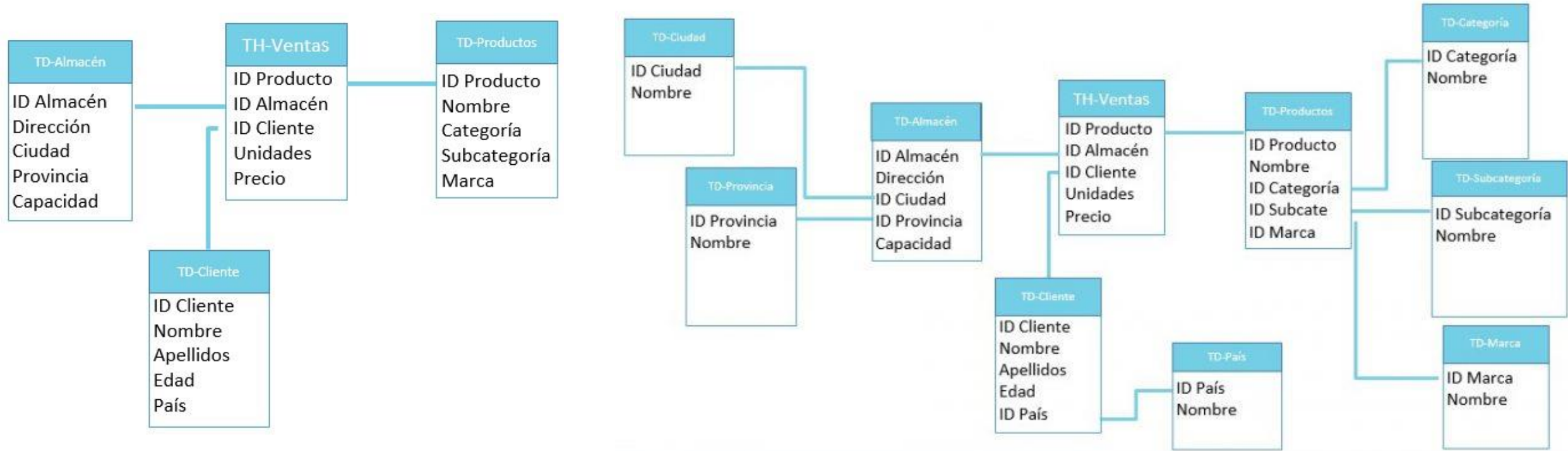
Requieren conocimiento de programación específico para su recolección, tratamiento, monitoreo y formalización

Bases de datos relacionales y datos estructurados

DIPLOMADO EN

 **PRODUCT
MANAGER**

ÁREA DE MARKETING



Bases de datos relacionales y datos estructurados

Dimensiones y métricas

Dimensiones		Métricas	
Productos		Número	Ventas
Producto 1		10	1.000
Producto 2		20	2.000
Producto 3		40	3.000
Producto 4		50	10.000
Producto 5		30	3.500
Producto 6		20	2.500
Producto 7		100	11.500

Bases de datos relacionales y datos estructurados

Dimensiones y métricas

Dimensiones

Métricas

Productos	Cadena	Número	Ventas
Producto 1	Cadena 1	5	600
Producto 1	Cadena 2	5	400
Producto 2	Cadena 1	15	1.400
Producto 2	Cadena 2	5	600
Producto 3	Cadena 1	40	3.000
Producto 4	Cadena 2	50	10.000
Producto 5	Cadena 1	30	3.500
Producto 6	Cadena 2	20	2.500
Producto 7	Cadena 1	75	9.000
Producto 7	Cadena 2	25	2.500

Bases de datos relacionales y datos estructurados

Diferencias y tratamientos

Productos	Cadena	Número	Ventas
Producto 1	Cadena 1	5	600
Producto 1	Cadena 2	5	400
Producto 2	Cadena 1	15	1.400
Producto 2	Cadena 2	5	600
Producto 3	Cadena 1	40	3.000



Productos	Cadena	Número	Ventas
Producto 1	Cadena 1	3	360
Producto 1	Cadena 1	2	240
Producto 1	Cadena 2	2	160
Producto 1	Cadena 2	2	160
Producto 1	Cadena 2	1	80
Producto 2	Cadena 1	15	1.400
Producto 2	Cadena 2	5	600
Producto 3	Cadena 1	10	750
Producto 3	Cadena 1	30	2.250



Bases de datos relacionales y datos estructurados

Ejemplo

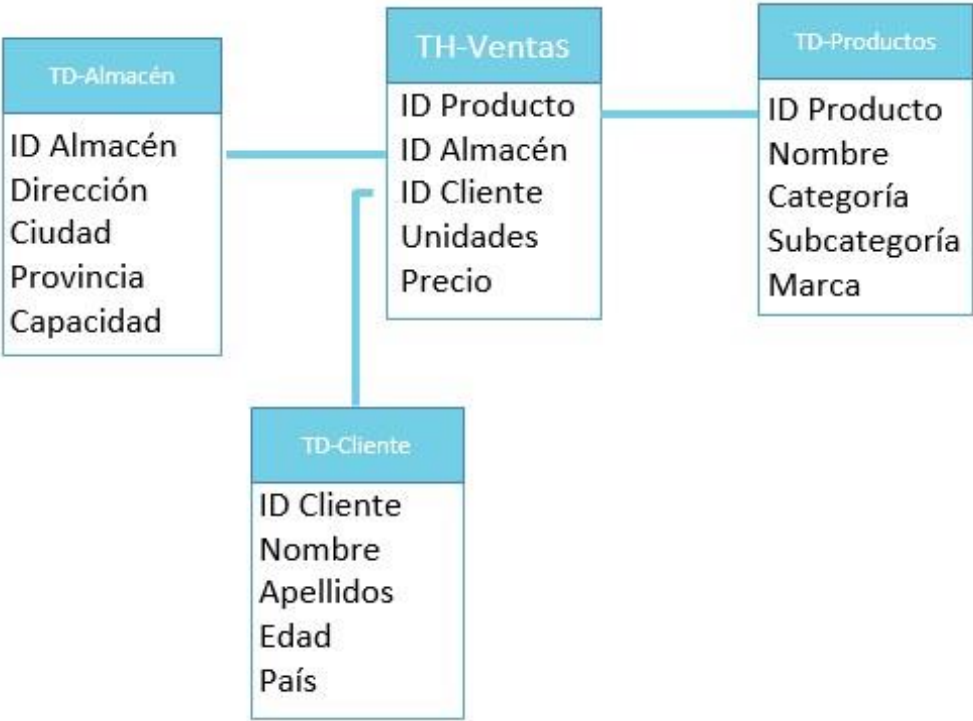


TD-Almacén

ID Almacén	Dirección	Ciudad	Provincia	Capacidad
234	Av. El Bosque	Santiago	Santiago	100
245	Independencia	Santiago	Santiago	20
235	Av. Ossa	Santiago	Santiago	30
453	Av. Tobalaba	Santiago	Santiago	500
246	Irarrázabal	Santiago	Santiago	20
124	Chile España	Santiago	Santiago	10
767	Walker Martinez	Santiago	Santiago	200

Bases de datos relacionales y datos estructurados

Ejemplo

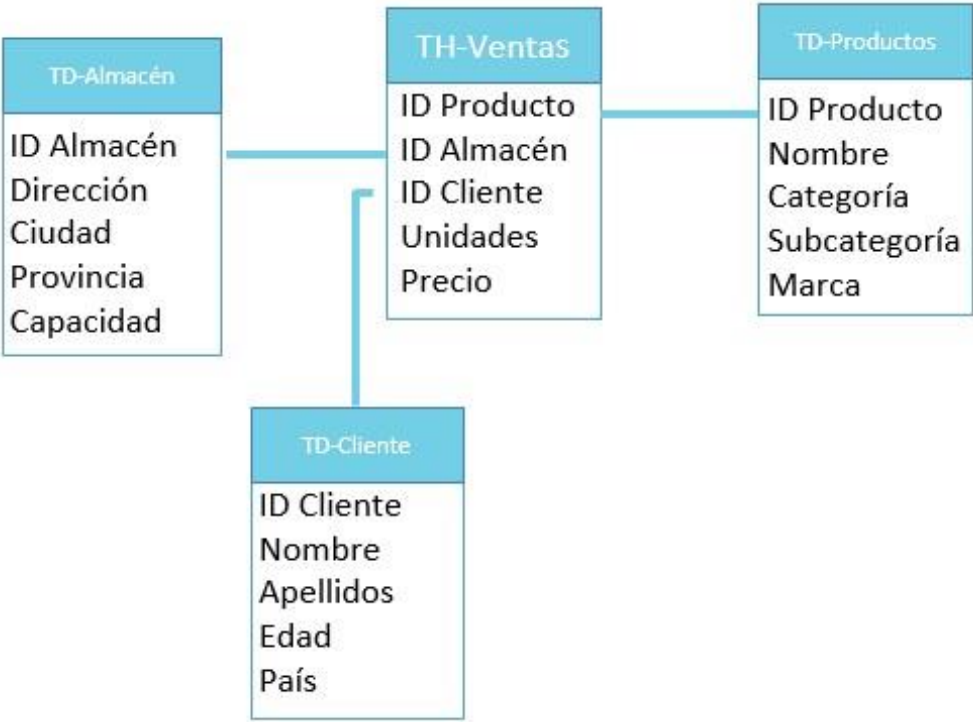


TD-Cliente

ID Cliente	Nombre	Apellidos	Edad	País
2343	Javier	Pérez	33	Chile
2455	Bernardita	González	34	Chile
2354	José	Gutiérrez	54	Chile
4533	Francisco	Pavlov	22	Chile
2462	Josefina	Montt	33	Chile
1241	María	Marín	55	Chile
7673	Flor	Henríquez	66	Chile

Bases de datos relacionales y datos estructurados

Ejemplo

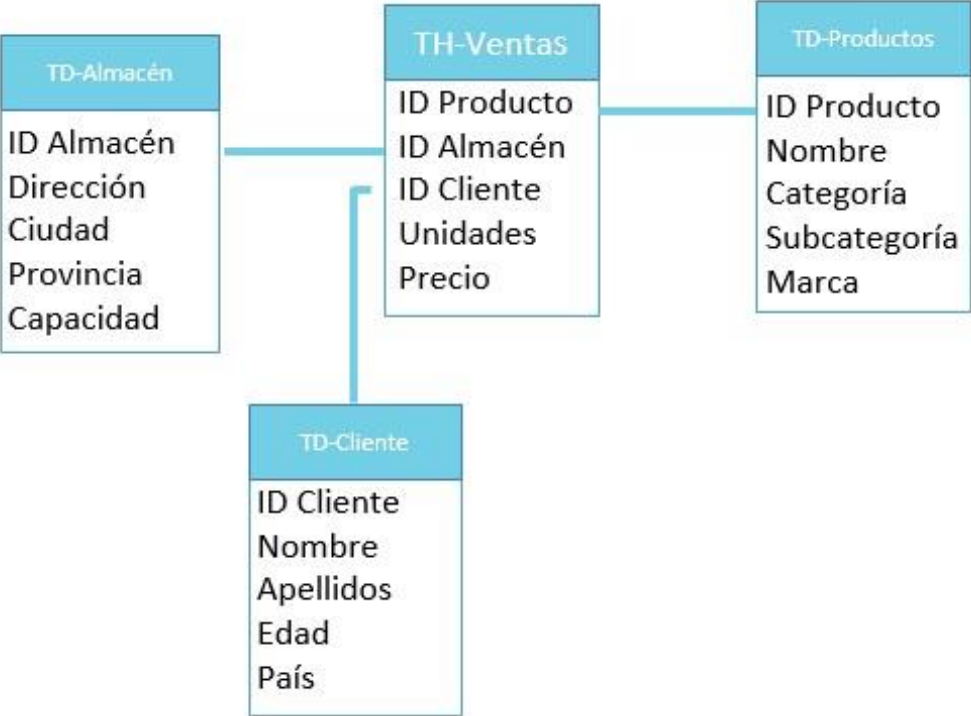


TD-Productos

ID Producto	Nombre	Categoría	Subcategoría	Marca
234321	Mostaza	Abarrotes	Salsa	Heinz
245512	Mermelada	Abarrotes	Untables	Watt's
235434	Jugo de Uva	Abarrotes	Jugos	CCU
453354	Durazno en conserva	Abarrotes	Conservas	Aconcagua
246245	Papas fritas	Congelados	Hortalizas	Minuto Verde
124145	Hamburguesa	Congelados	Congelados	NotCo
767345	Manjar	Abarrotes	Untables	Nestle

Bases de datos relacionales y datos estructurados

Ejemplo



TH-Ventas

ID Producto	ID Almacén	ID Cliente	Unidades	Precio
234321	234	2343	30	1.200
245512	245	2455	120	1.300
235434	235	2354	400	2.000
453354	453	4533	650	4.500
246245	246	2462	300	2.000
124145	124	1241	494	5.400
767345	767	7673	600	6.000

Flujo de datos

DIPLOMADO EN

**PRODUCT
MANAGER**

ÁREA DE MARKETING

TI

Recolección

Se reúne data confiable desde distintas fuentes



Almacenamiento

Se almacena la data de manera apropiada en un data warehouse, data lake o data mart



Preparación

Se evalúa la completitud, precisión y otras métricas de calidad



Procesamiento

Se verifica, transforma y organiza la data



Análisis

Se analiza e interpreta la data en el contexto de preguntas específicas del negocio

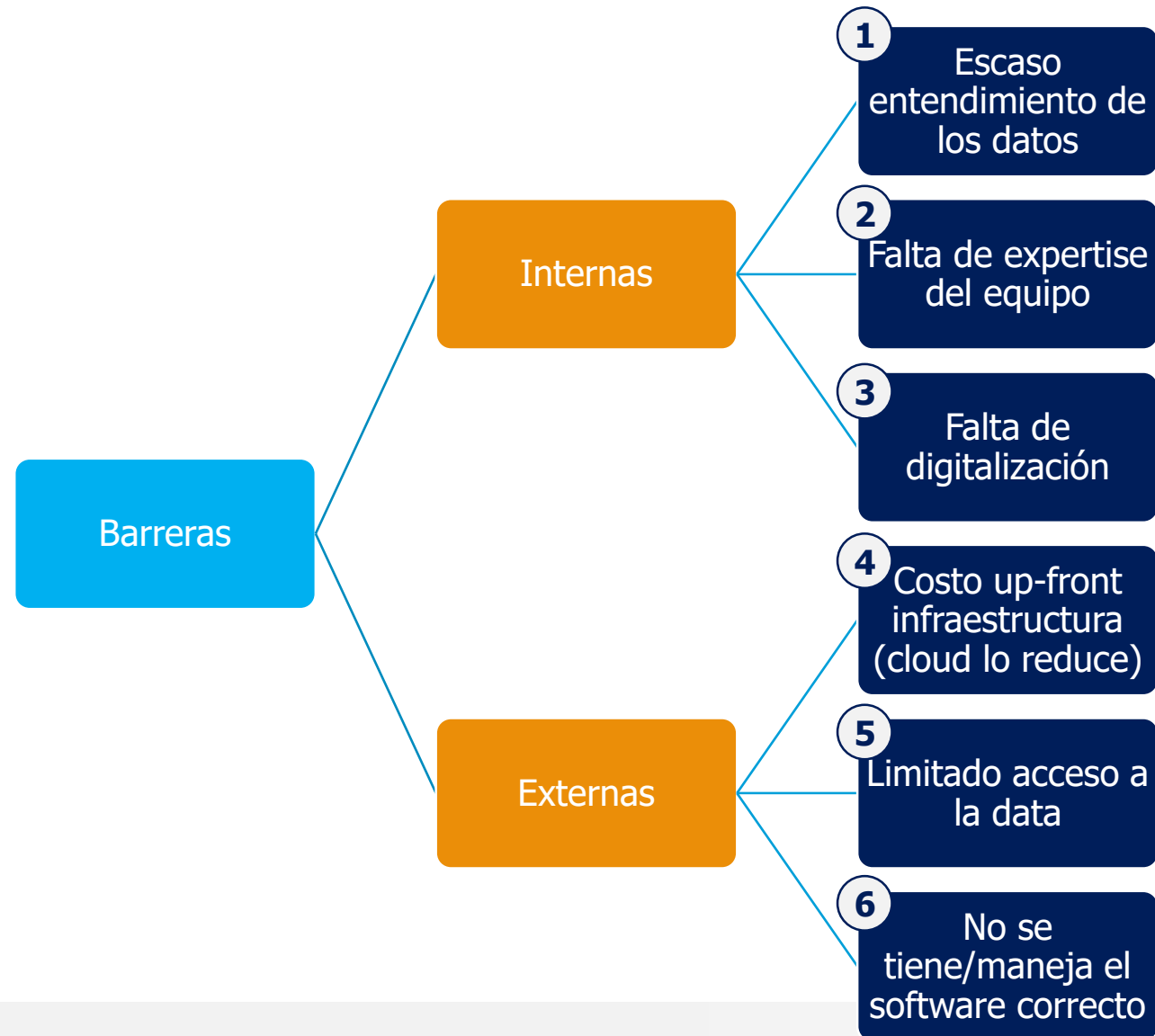


Presentación

Se presenta la data al usuario de una manera que haga sentido, en forma de tablas, gráficos y otras visualizaciones



Barreras de adquisición y tratamiento de datos



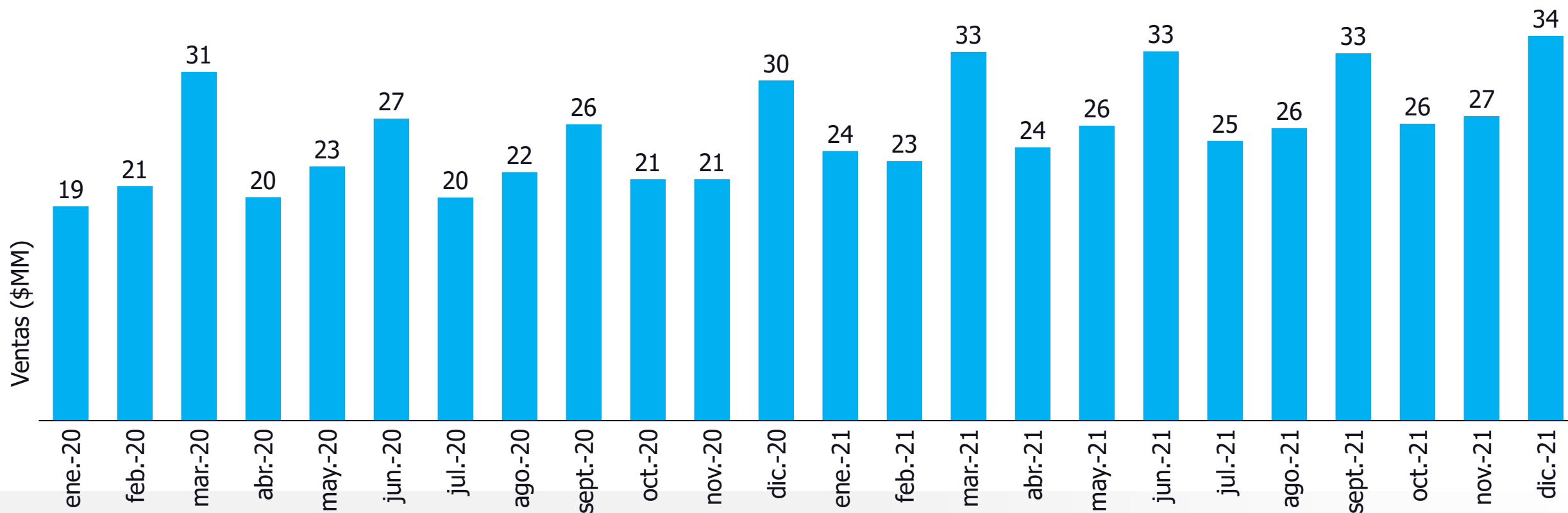
Barreras de adquisición y tratamiento de datos Internas

1

Escaso entendimiento de los datos

PRODUCT
MANAGER

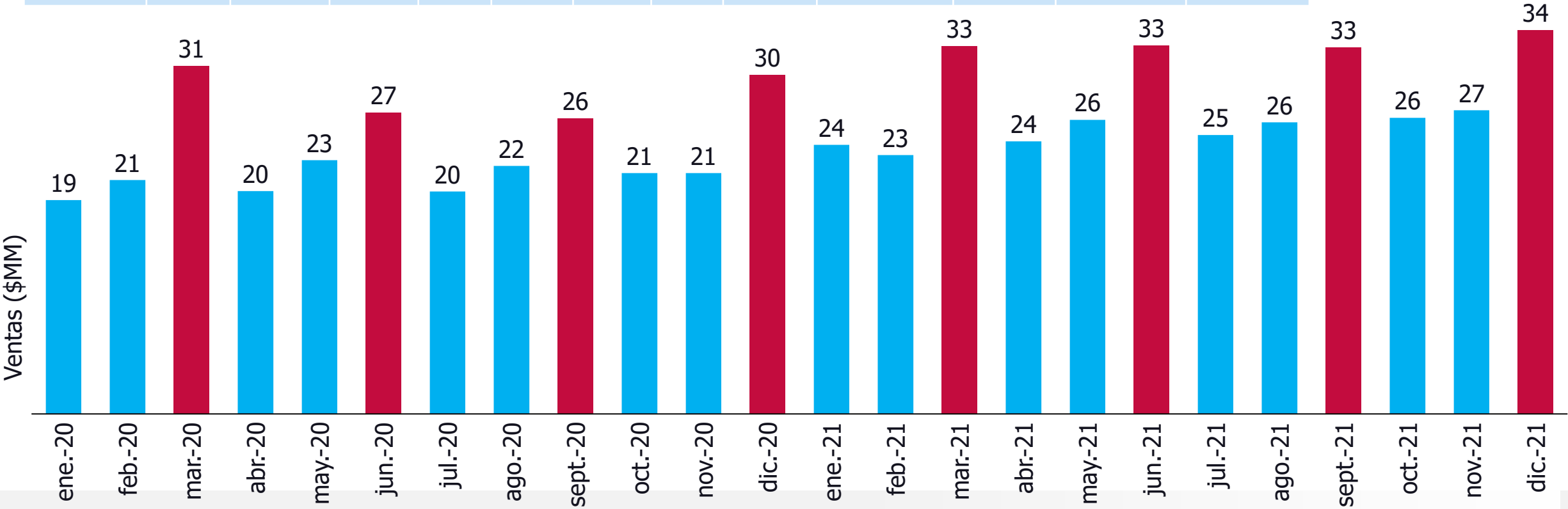
ÁREA DE MARKETING



Barreras de adquisición y tratamiento de datos Internas

1 Escaso entendimiento de los datos

	enero	febrero	marzo	abril	mayo	junio	julio	agosto	septiembre	octubre	noviembre	diciembre
# semanas	4	4	5	4	4	5	4	4	5	4	4	5



Barreras de adquisición y tratamiento de datos Internas

2

Falta de expertise del equipo

- No conocer el origen de los datos
- No aplicar el modelo correcto

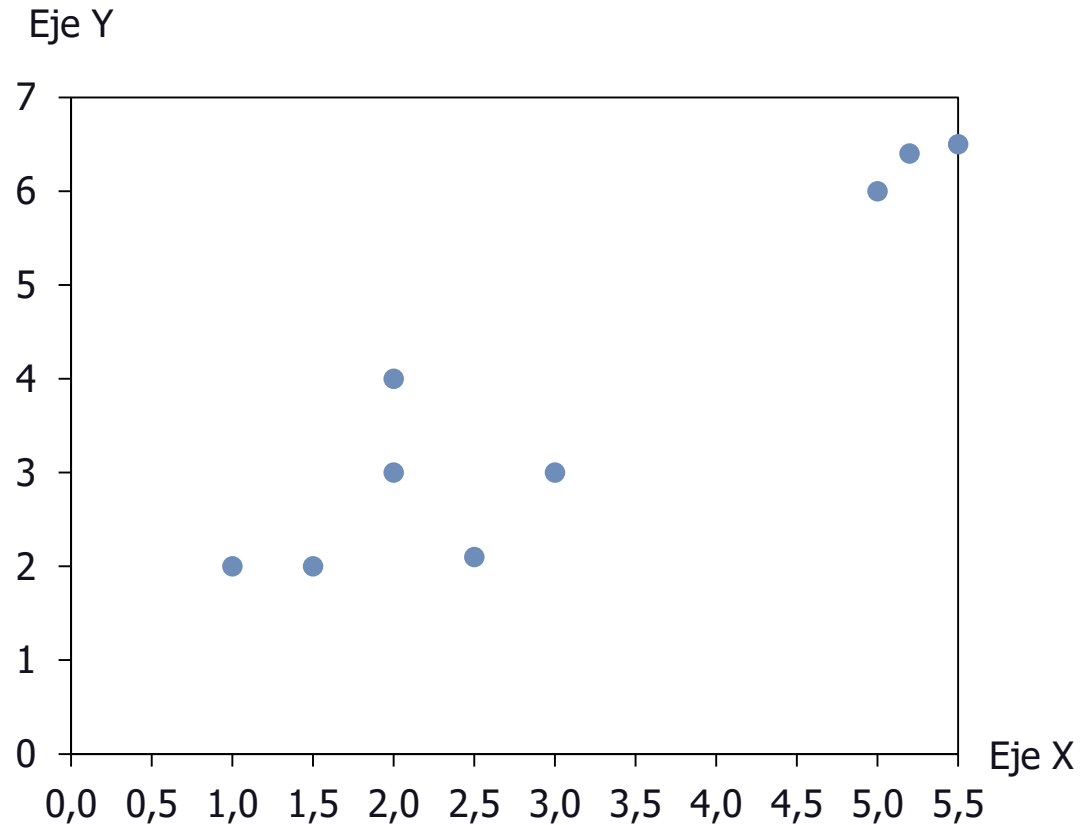


Barreras de adquisición y tratamiento de datos Internas

2

Falta de expertise del equipo

- No conocer el origen de los datos
- No aplicar el modelo correcto

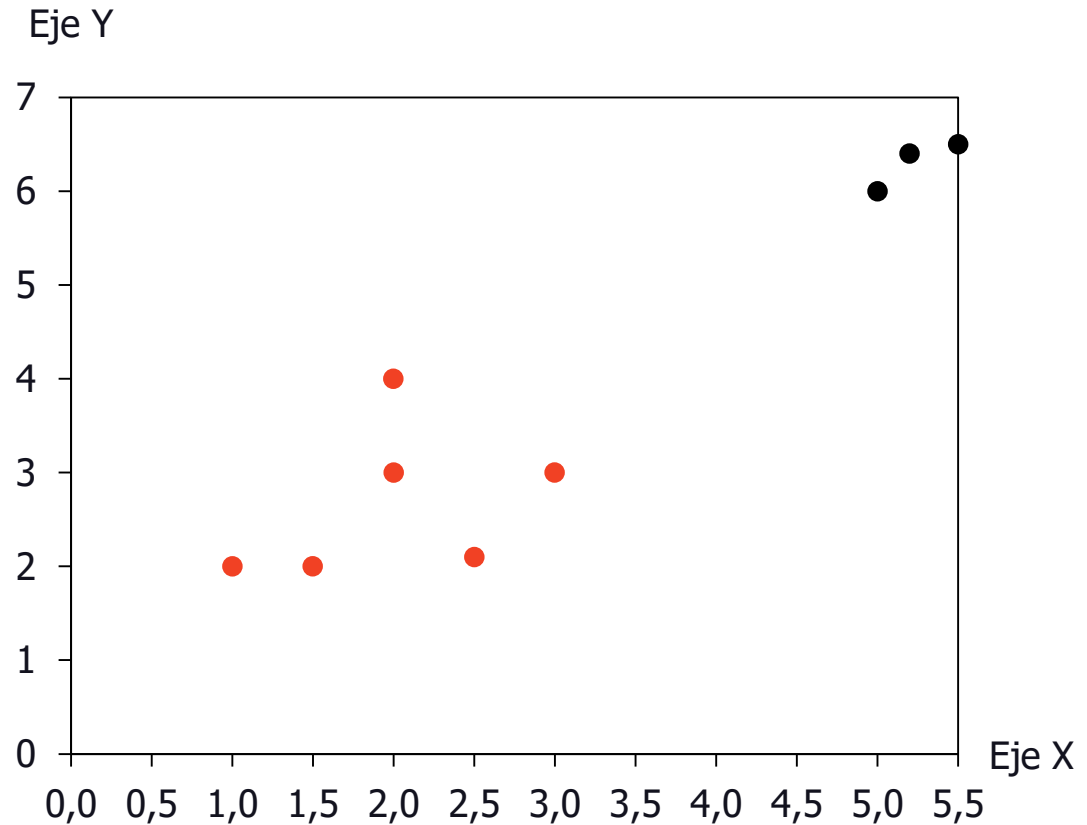


Barreras de adquisición y tratamiento de datos Internas

2

Falta de expertise del equipo

- No conocer el origen de los datos
- No aplicar el modelo correcto



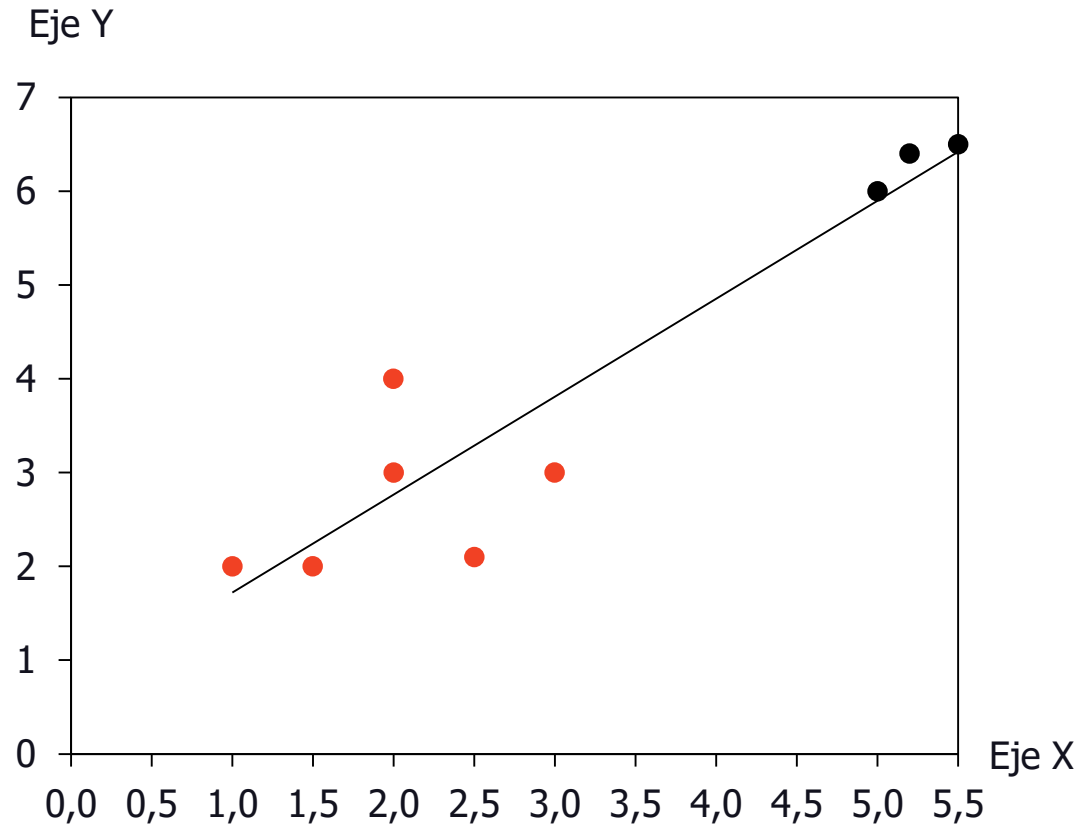
Barreras de adquisición y tratamiento de datos Internas

2

Falta de expertise del equipo

- No conocer el origen de los datos
- No aplicar el modelo correcto

$$E_d = \frac{\frac{\Delta Q_d}{Q_d}}{\frac{\Delta P}{P}}$$



DIPLOMADO EN

**PRODUCT
MANAGER**

ÁREA DE MARKETING

Barreras de adquisición y tratamiento de datos Internas

3

Falta de digitalización

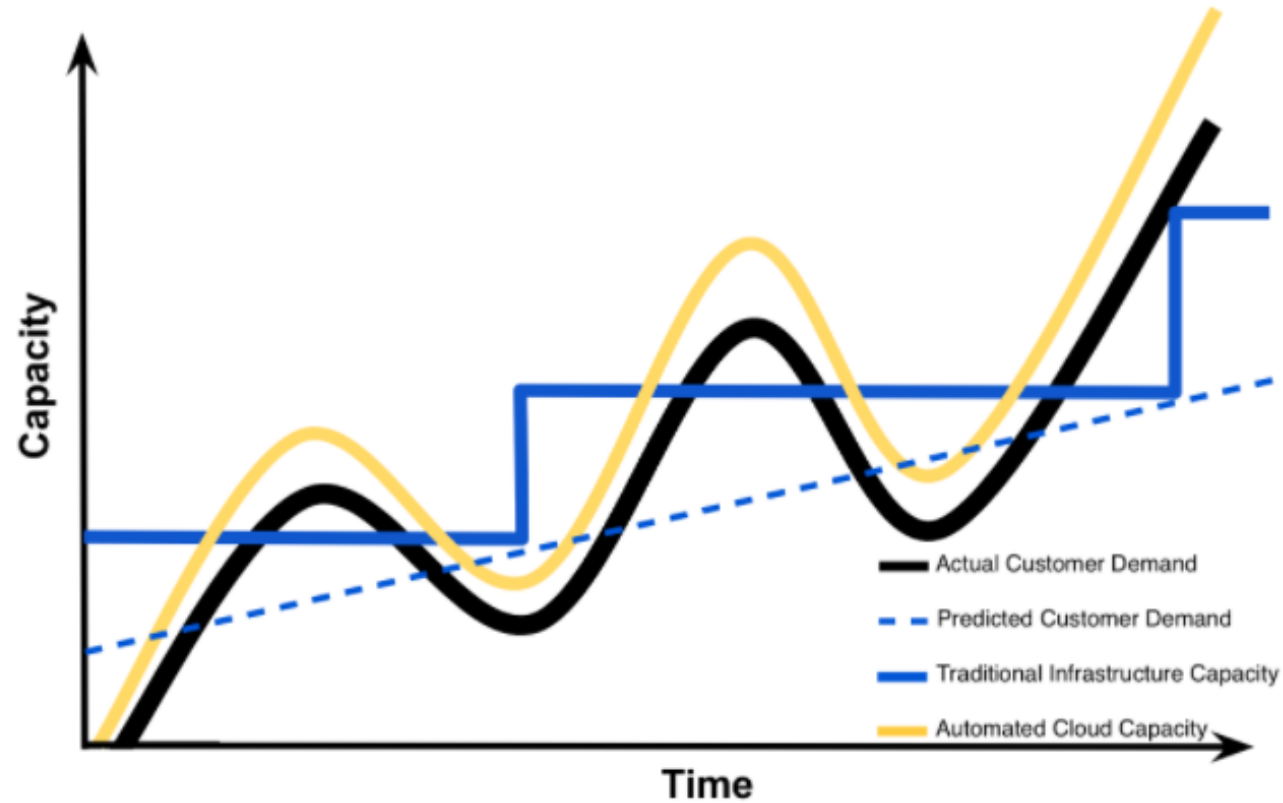
- No existe la infraestructura para trabajar con la data disponible
- La data que existe se reúne de manera manual



Barreras de adquisición y tratamiento de datos Externas

4

Costo up-front
infraestructura
(cloud lo reduce)

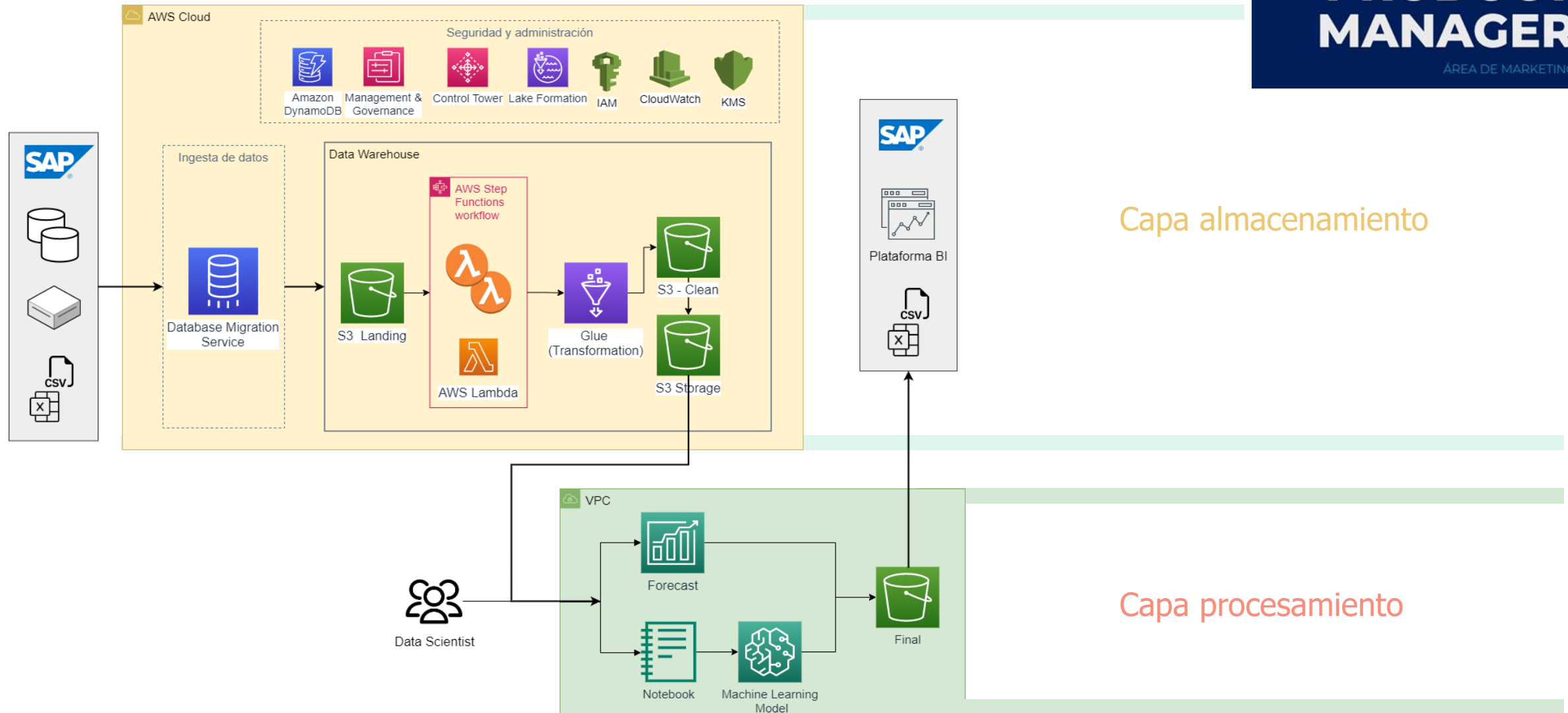


DIPLOMADO EN

**PRODUCT
MANAGER**

ÁREA DE MARKETING

Un paréntesis...



Barreras de adquisición y tratamiento de datos Externas

5

Limitado acceso a la data

- No se tienen los recursos para adquirir información clave de mercado
- Se desconoce la rentabilidad de hacerlo



Barreras de adquisición y tratamiento de datos Externas

6

No se
tiene/maneja el
software correcto

- Plataforma de deployment no conversa con el software de análisis
- Esfuerzos superiores en integrar flujo de data science en el negocio



Bases relacionales y datos estructurados

Ejercicio #1



En la Empresa XYZ usted debe definir, si existe, la data que utilizará para resolver las siguientes interrogantes:

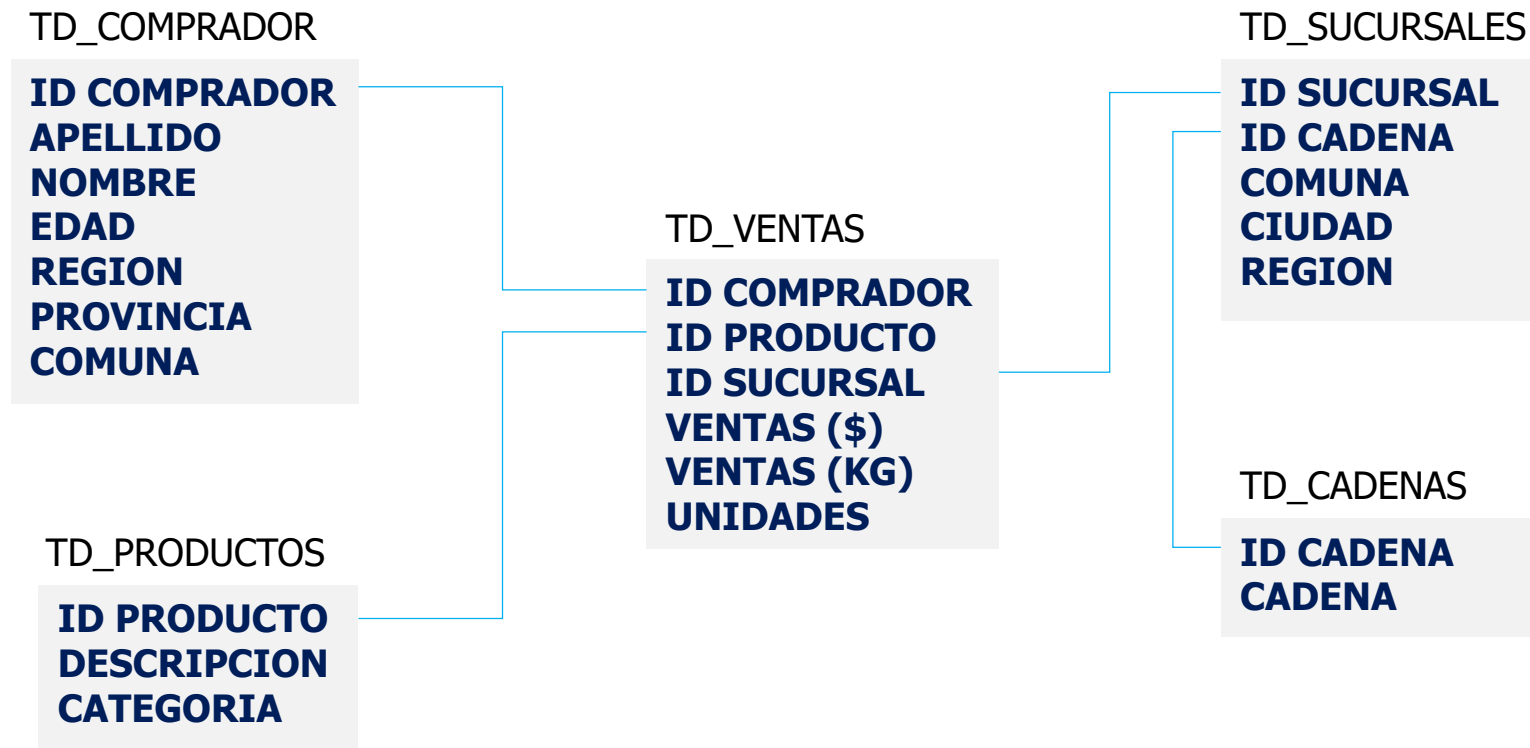
1. ¿Cuál es la edad promedio del comprador de mermeladas?
2. ¿Cuál ha sido el peak de ventas de los productos untables?
3. ¿En qué región conviene vender los productos?
4. ¿Qué categoría tiene mayor Precio Promedio?
5. ¿Cuál es la rentabilidad de las conservas? Si es el caso, ¿Desde dónde puedo obtener la información faltante?

Bases relacionales y datos estructurados

Ejercicio #1





A continuación se muestra un esquema con los datos que la empresa XYZ maneja actualmente:







Gracias

 Cristian León Meza

 +569 5696 2718

 cristian.leon.me@usach.cl

 <https://www.linkedin.com/in/cristian-león>