



ESCUELA DE INGENIERÍA
FACULTAD DE INGENIERÍA

EDUCACIÓN
PROFESIONAL

Diplomado en Big Data y Ciencia de Datos
Ciencia de Datos y sus Aplicaciones

Clase 04:

Aplicaciones en la Industria y Fuga de Clientes

Roberto González



regonzar@uc.cl





Clase 04: Aplicaciones en la Industria y Fuga de Clientes

INDUSTRIA GUIADA POR DATOS

Data Scientist

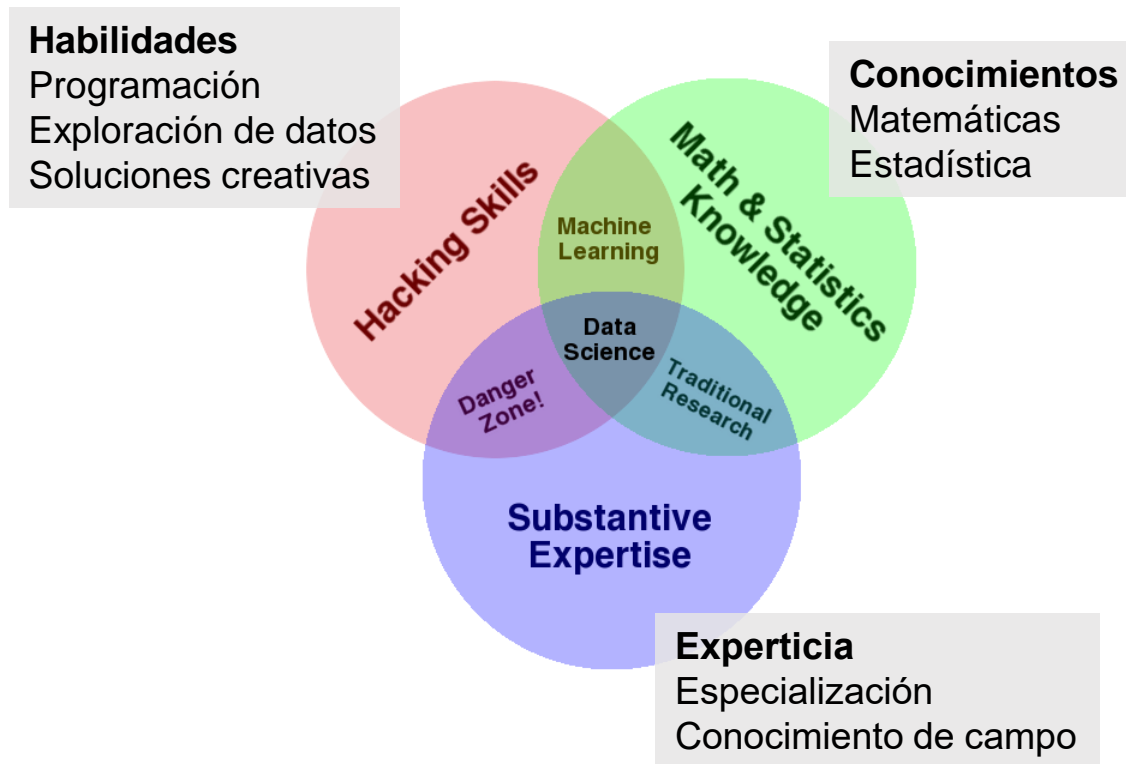
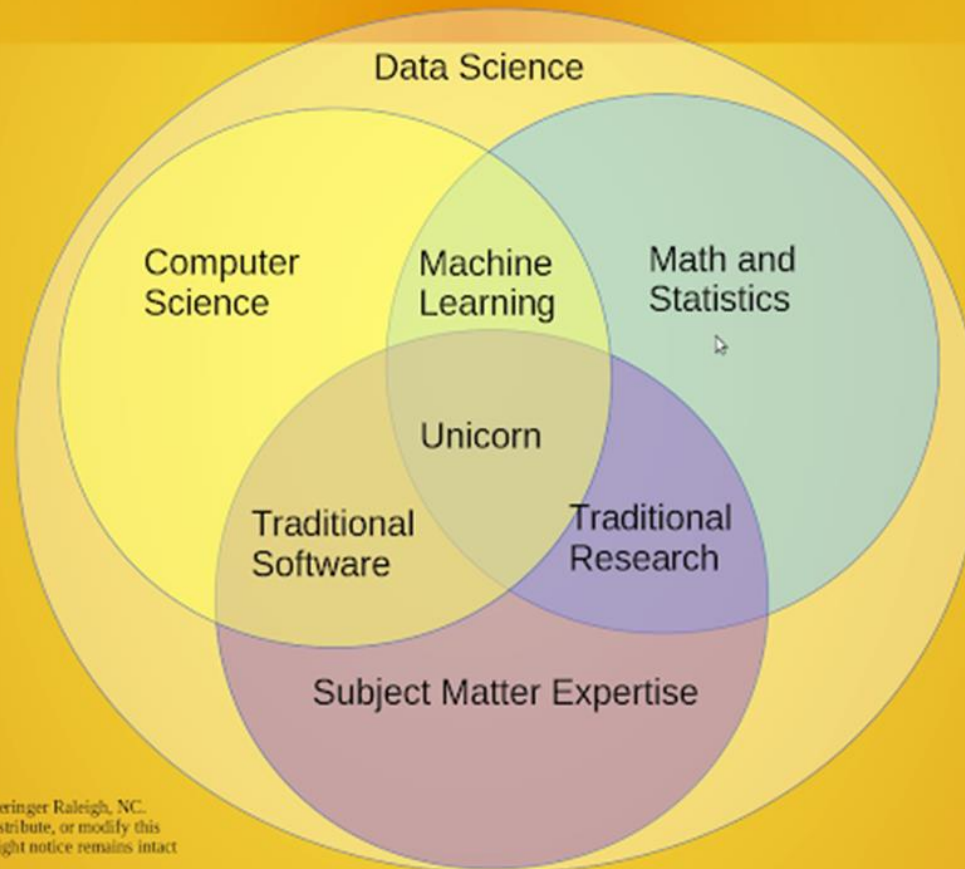


Diagrama de Venn para Data Science
Drew Conway (2010)

Data Scientist

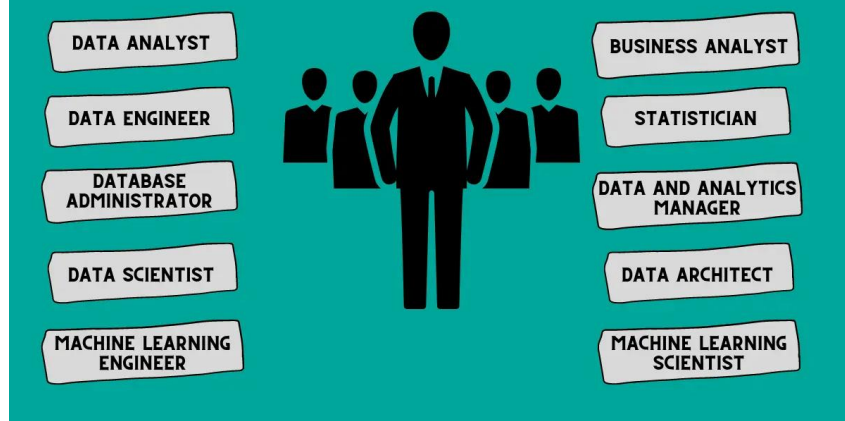
Data Science Venn Diagram v2.0



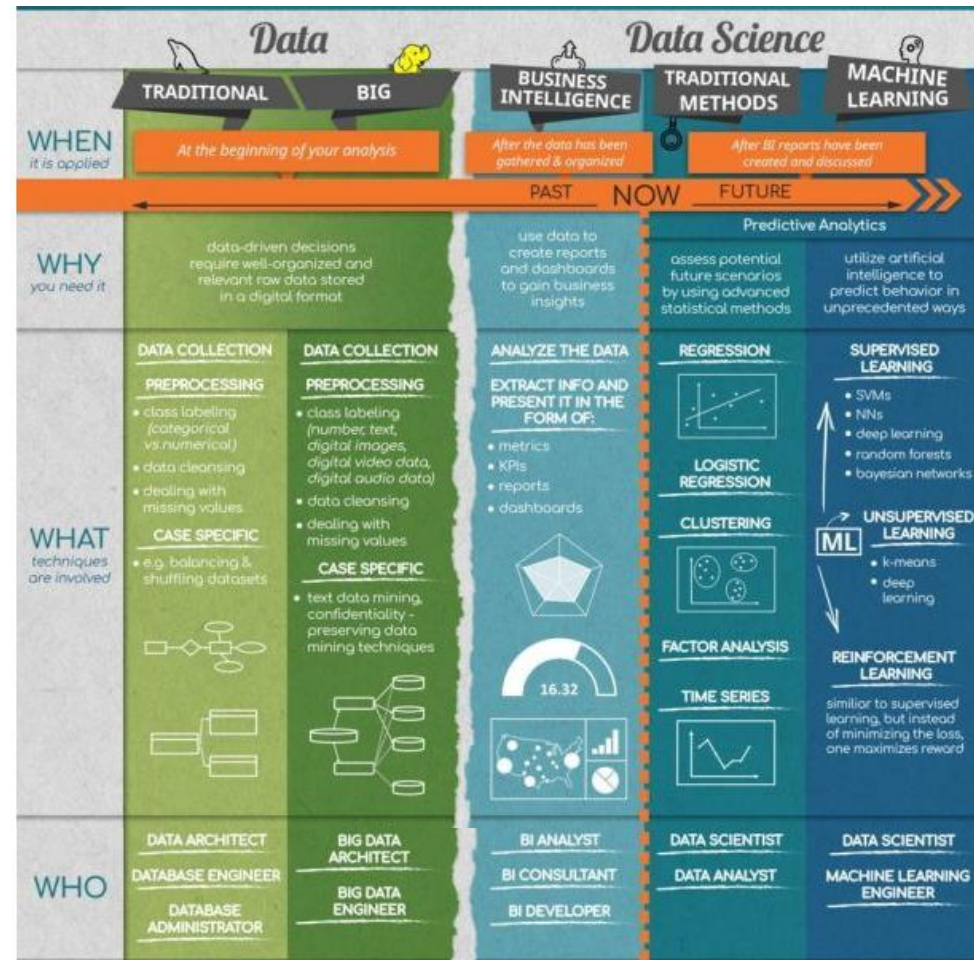
Copyright © 2014 by Steven Geringer Raleigh, NC.
Permission is granted to use, distribute, or modify this
image, provided that this copyright notice remains intact

Roles

Data Science Job Roles



	Data Analyst	Data Scientist	Data Engineer	Machine Learning Researcher	Machine Learning Engineer	Full Stack
Sales	Growth Analyst	Growth Data Scientist	Growth Data Engineer			Growth Data Scientist - Full Stack
Marketing						
Strategy						
Operations	BI Analyst	Operations Data Scientist	BI/Operations Data Engineer			BI/Operations Data Scientist - Full Stack
Finance						
Product	Product Analyst	Product Data Scientist	Product Data Engineer			Product Data Scientist - Full Stack
R&D		Data Scientist	Data Engineer	ML Researcher	ML Engineer	Data Scientist - Full Stack
Engineering						



Medium.com



Tipos de datos

- Los datos son el punto de partida para todo análisis.
- **Estructurados:** Están altamente organizados.
Se almacenan en una base de datos relacional.



Año	PIB (\$ Millones)	Consumo Eléctrico (GWh)
1993	32.559.292	21.011,3
1994	34.416.724	22.730,7
1995	38.028.591	24.910,2
1996	40.831.596	27.969,0
1997	43.526.546	30.351,5
1998	44.944.340	33.015,8
1999	44.616.349	35.921,3
2000	46.605.199	38.867,4

Tipos de datos

- **No estructurados:** Son datos crudos y no están organizados. Deben ser procesados y transformados para luego ser almacenados en una base de datos.



```
"sources" : [ {  
  "id" : "example_ID_0001",  
  "name" : "leicester_lab",  
  "urls" : [ "http://www.le.ac.uk/genetics" ],  
  "contacts" : [ {  
    "name" : "Owen Lancaster ",  
    "email" : "ejh@leicester.ac.uk",  
    "db_xrefs" : [ {  
      "accession" : "123456",  
      "source" : "orcid"  
    } ]  
  } ]  
} ]
```







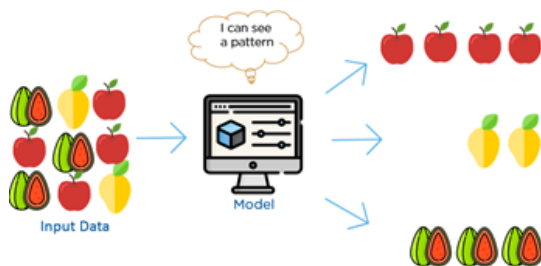
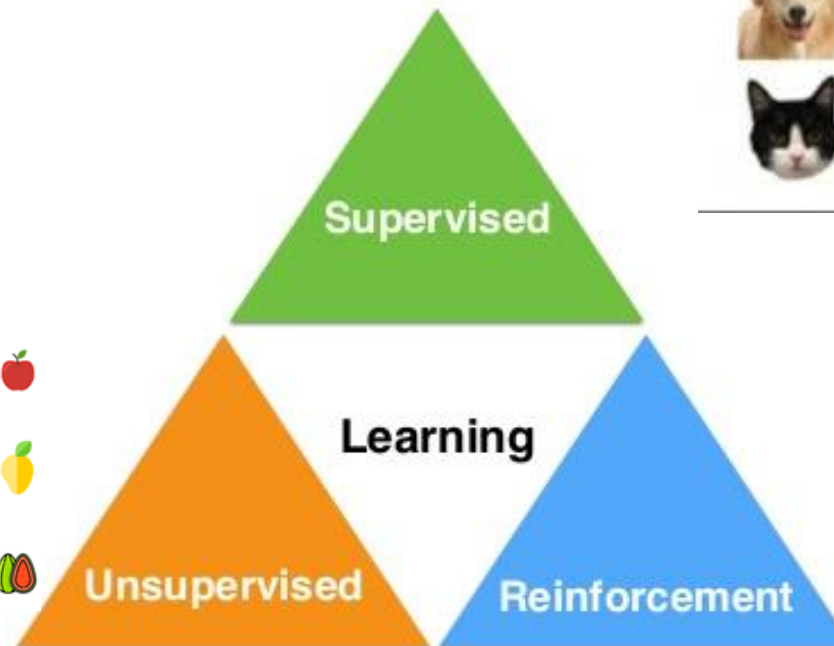
Clase 04: Aplicaciones en la Industria y Fuga de Clientes

MACHINE LEARNING

Tipos

- Labeled data
- Direct feedback
- Predict outcome/future

Input	Label	Prediction
	CAT	
	NOT CAT	
	CAT	
		 ?

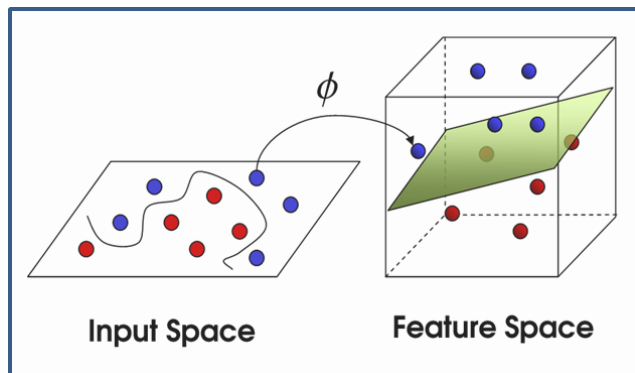


- No labels
- No feedback
- "Find hidden structure"

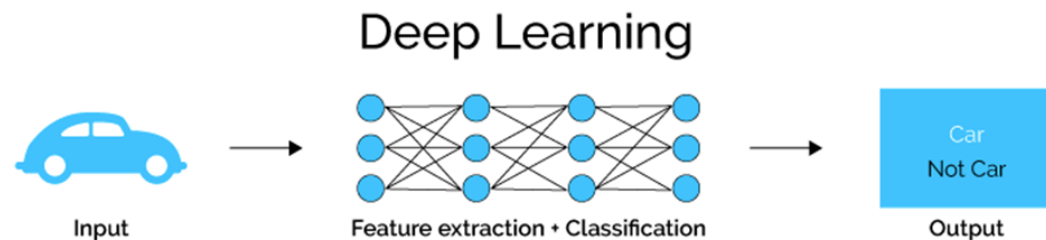
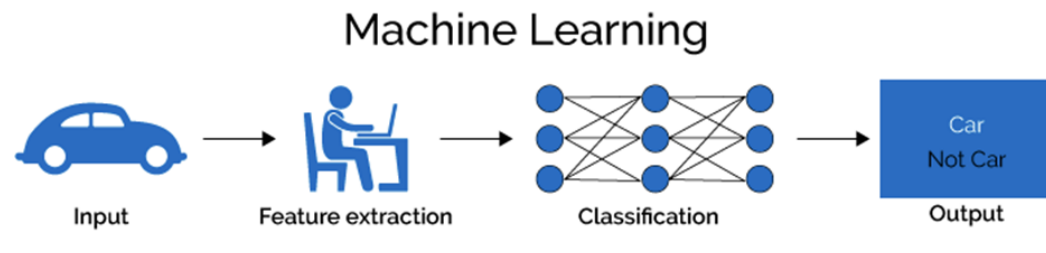
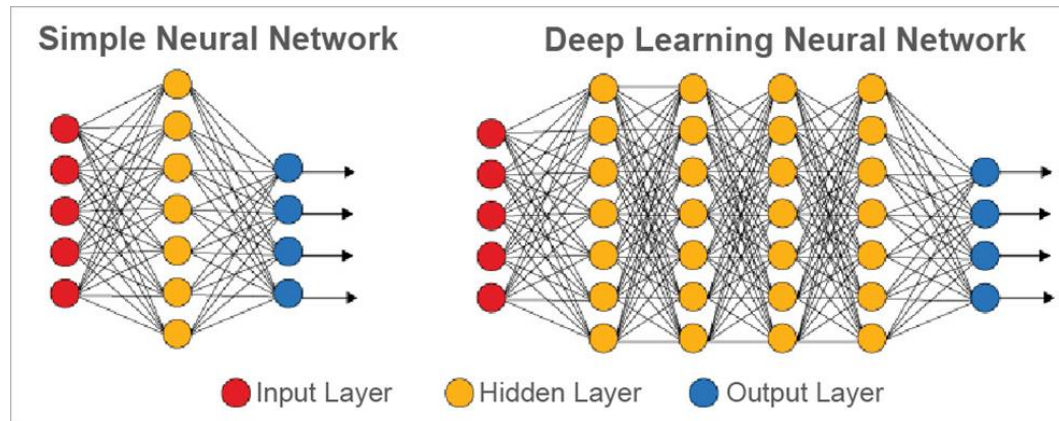


- Decision process
- Reward system
- Learn series of actions

SVM – ML - DL



Support Vector Machine :
Modelos+Algoritmos para
mapear datos en espacios
de mayor dimensionalidad
(feature space) para poder
separarlos según un
hiperplano. Sirve para
clasificación y regresión.





Clase 04: Aplicaciones en la Industria y Fuga de Clientes

APLICACIONES

Fintech <https://www.youtube.com/watch?v=MZrs7zQU068>



La penetración digital está elevando las expectativas de los clientes

67%

de los clientes dicen que su nivel de buenas experiencias es más alto que nunca



72%

de los compradores esperan que los vendedores personalicen la experiencia a sus necesidades personales

80%

de los clientes dicen que la experiencia de una empresa es tan importante como sus productos y servicios

59%

de los clientes dicen que las empresas necesitan experiencias digitales de vanguardia para mantener su negocio. El 59% también dice que están abiertos a que las empresas usen la inteligencia artificial (AI) para mejorar su experiencia

Fuente: EY

www.educacionprofesional.ing.uc.cl

El desarrollo de soluciones de Data & Analytics son una mezcla de múltiples tecnologías que requieren diferentes capacidades

Canales



Resultados



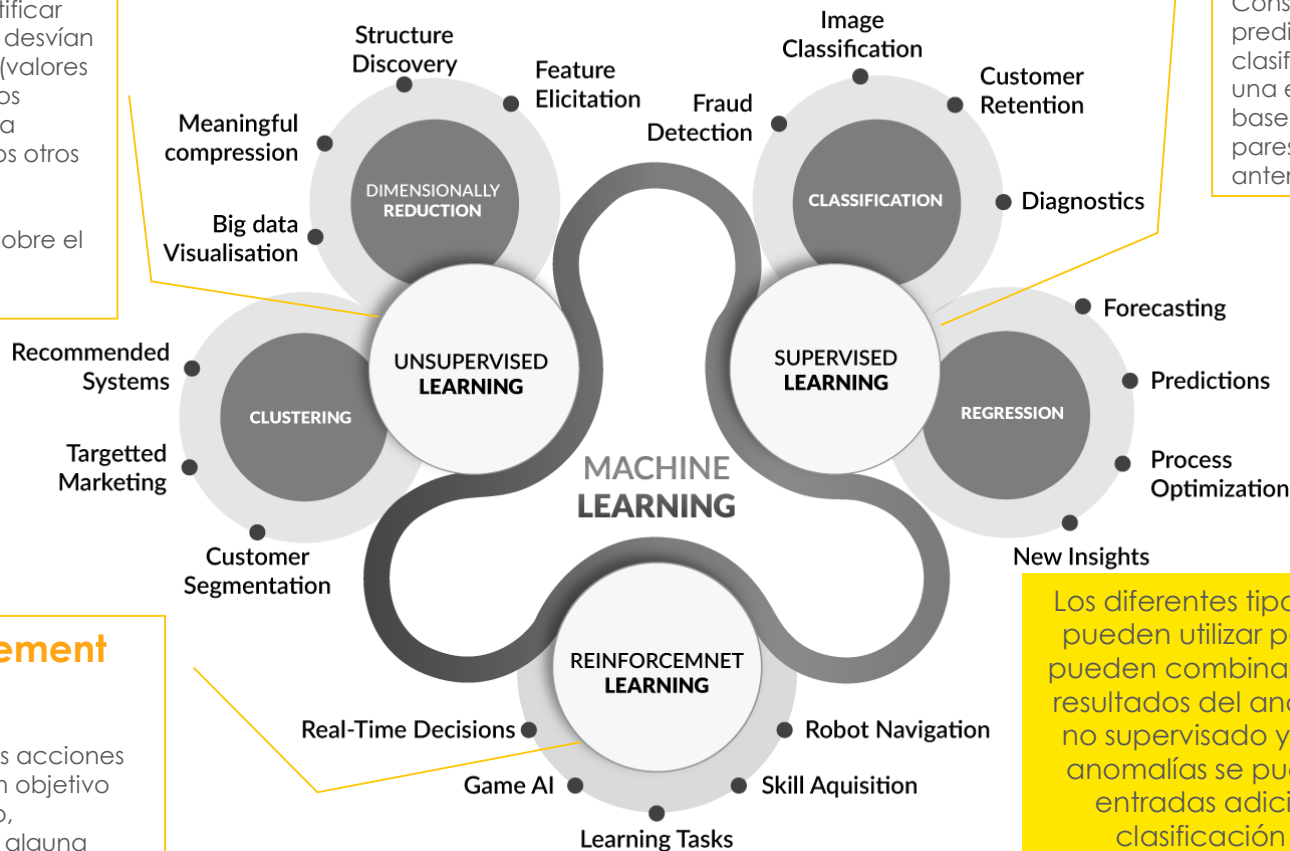
Enfoque para los distintos tipos de aprendizaje

Unsupervised learning

Los modelos pueden agrupar datos no estructurados y/o pueden identificar casos que se desvían fuertemente (valores atípicos) de los patrones de la mayoría de los otros casos. No se proporciona información sobre el resultado al algoritmo.

Supervised learning

Construyendo modelos predictivos o de clasificación que mapean una entrada a una salida en base a observaciones de pares de entrada-salida anteriores.



Reinforcement learning

Determinar las acciones para lograr un objetivo en un entorno, maximizando alguna recompensa.

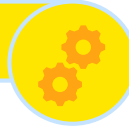
Los diferentes tipos de enfoques se pueden utilizar por separado o se pueden combinar. Por ejemplo: Los resultados del análisis de clustering no supervisado y la detección de anomalías se pueden usar como entradas adicionales en una clasificación supervisada.



Aplicaciones de Data Science y Machine Learning

- Mantenimiento Predictivo
- Propensión de compras
- Estimación de demanda
- Optimización de procesos

Manufactura



- Planificación de inventario predictivo
- Motores de recomendación
- Ventas y marketing multicanal
- Segmentación del mercado
- Retorno de la inversión del cliente y lifetime value

Retail



- Programación del tráfico aéreo
- Pricing dinámico
- Análisis de interacción
- Resolución de quejas de los clientes
- Análisis de patrones de tráfico y análisis de congestión

Travel & Hospitality



- Análisis de riesgos
- Segmentación de clientes
- Venta cruzada y up-selling
- Gestión de campañas de ventas y marketing
- Evaluación de la solvencia crediticia

Servicios Financieros





Aplicaciones de Data Science y Machine Learning

- Alertas y diagnósticos a partir de datos de pacientes en tiempo real
- Identificación de enfermedades y estratificación de riesgos
- Optimización del triaje de pacientes

Salud



- Análisis del uso de la energía
- Emisiones de Carbono
- Gestión de smart grid
- Optimización de la demanda y la oferta de energía

Energy & Utilities



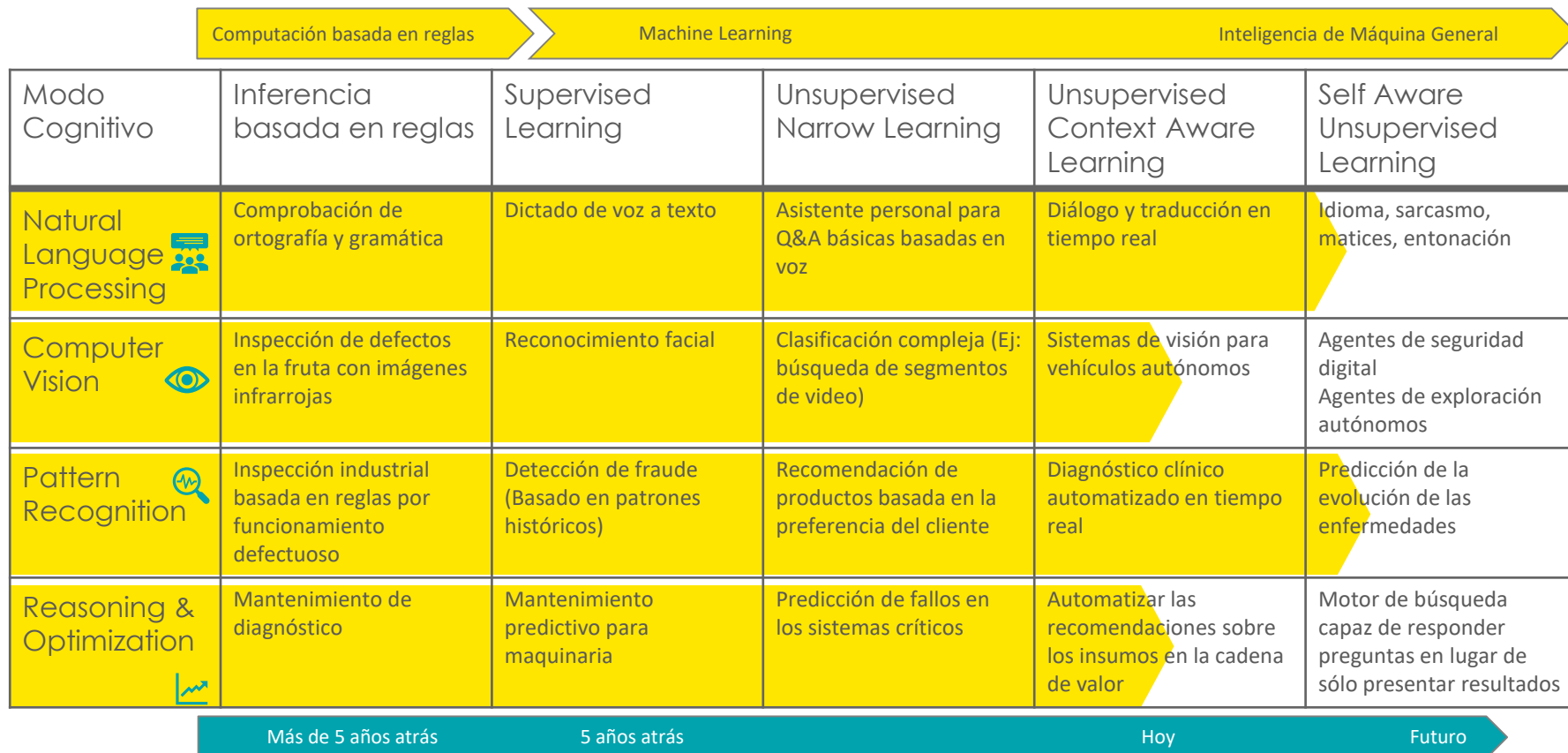
- Customer Churn Analysis
- Detección de Fraude
- Clickstream Analysis
- Gestión de Red / Optimización
- Gestión de la capacidad instalada

Telecomunicaciones





Casos de Uso de Machine Learning



Un ejemplo: Surgen nuevos modelos de servicios bancarios

	Banco como experiencia	Banco como marketplace	Banco como service provider	Banco como facilitador	Banco como productor
Servicios Clave	Proporcionar una experiencia superior al cliente, para atraer grandes volúmenes de nuevos clientes y aumentar la lealtad de los ya existentes.	Crear nuevas propuestas de valor, como la ampliación de la banca con productos y servicios no bancarios.	Competir con otros proveedores. Un banco como proveedor de servicios es como un proveedor de outsourcing de TI pero con capacidad bancaria.	Proporcionar una plataforma que permita la innovación colaborativa. Los conocimientos adquiridos permitirán desarrollar servicios bancarios y no bancarios específicos.	Proporcionar operaciones y tecnología para sus clientes, ya sea como una propuesta bancaria completa o para productos o mercados específicos sobre la base de una estrategia white label.
Rol de Analytics y AI	<ul style="list-style-type: none">▶ Mejora la interacción con el cliente▶ Simplificación de tareas▶ Personalización del servicio	<ul style="list-style-type: none">▶ Recomendaciones personalizadas▶ Gestión de leads▶ Ofertas a la medida▶ Relación entre producto y cliente	<ul style="list-style-type: none">▶ Fraudes▶ Garantizar la fiabilidad y la calidad del servicio▶ Autenticación e identidad▶ Sistemas de reputación▶ Colocación de créditos	<ul style="list-style-type: none">▶ Gestión de las finanzas personales▶ Análisis de riesgos▶ Perspectivas del mercado y la competencia▶ Análisis de tendencias	<ul style="list-style-type: none">▶ Operaciones automatizadas avanzadas▶ Seguridad mejorada en un entorno abierto



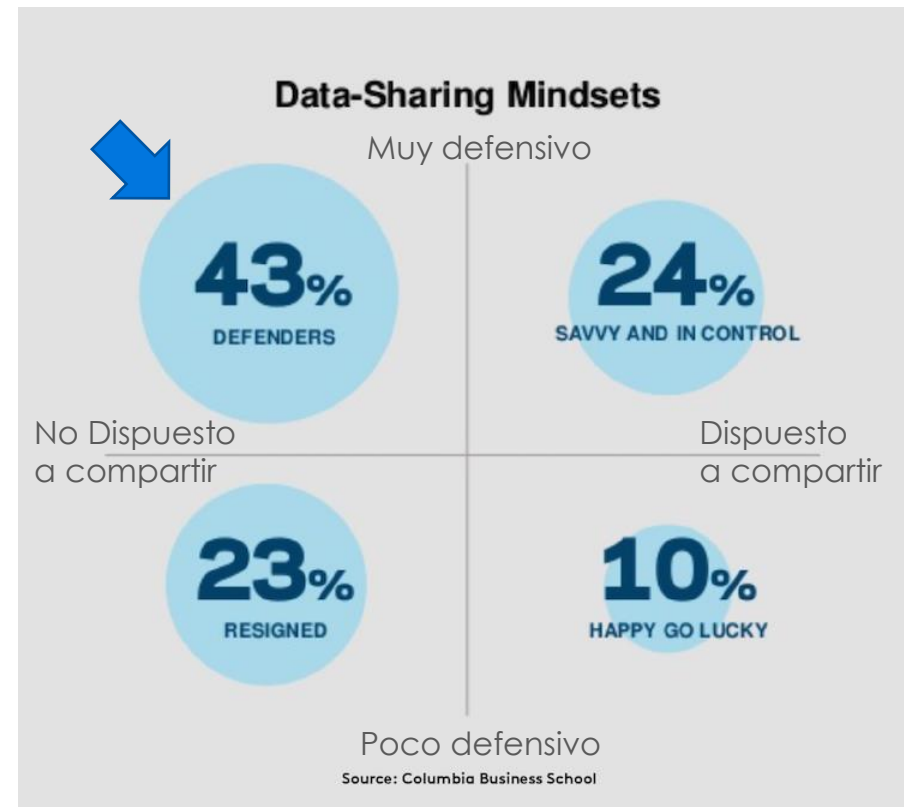
Pregunta: "¿Cuál de los siguientes factores consideras que son las principales limitaciones para la adopción de Data Science y Machine Learning?"

- 1) Necesidad de una cultura innovadora
- 2) La complejidad técnica de las soluciones analíticas
- 3) La falta de confianza hacia las soluciones analíticas
- 4) Explicabilidad y brecha de capacidades
- 5) Calidad/cantidad de los datos

Principales retos

- Confianza en las soluciones analíticas
- Gobernanza
- Auditabilidad y control
- Disponibilidad de datos significativos
- La explicabilidad de los resultados
- Privacidad de los datos

Davir Rogers, CBS. 2015 (8k consumers US,UK,Fr,Ca)



THE DATA-SHARING MINDSETS

Not all consumers respond to marketers' efforts to collect their personal data the same way. The quadrants below profile the ways consumers perceive and act on sharing their data:

DEFENDER

High Defense, Not Happy To Share

Millennials	36%
Gen X	41%
Boomers	48%
Silent Gen	46%

RESIGNED

Low Defense, Not Happy To Share

Millennials	20%
Gen X	22%
Boomers	24%
Silent Gen	29%

SAVVY AND IN CONTROL

High Defense, Happy To Share

Millennials	34%
Gen X	26%
Boomers	18%
Silent Gen	12%

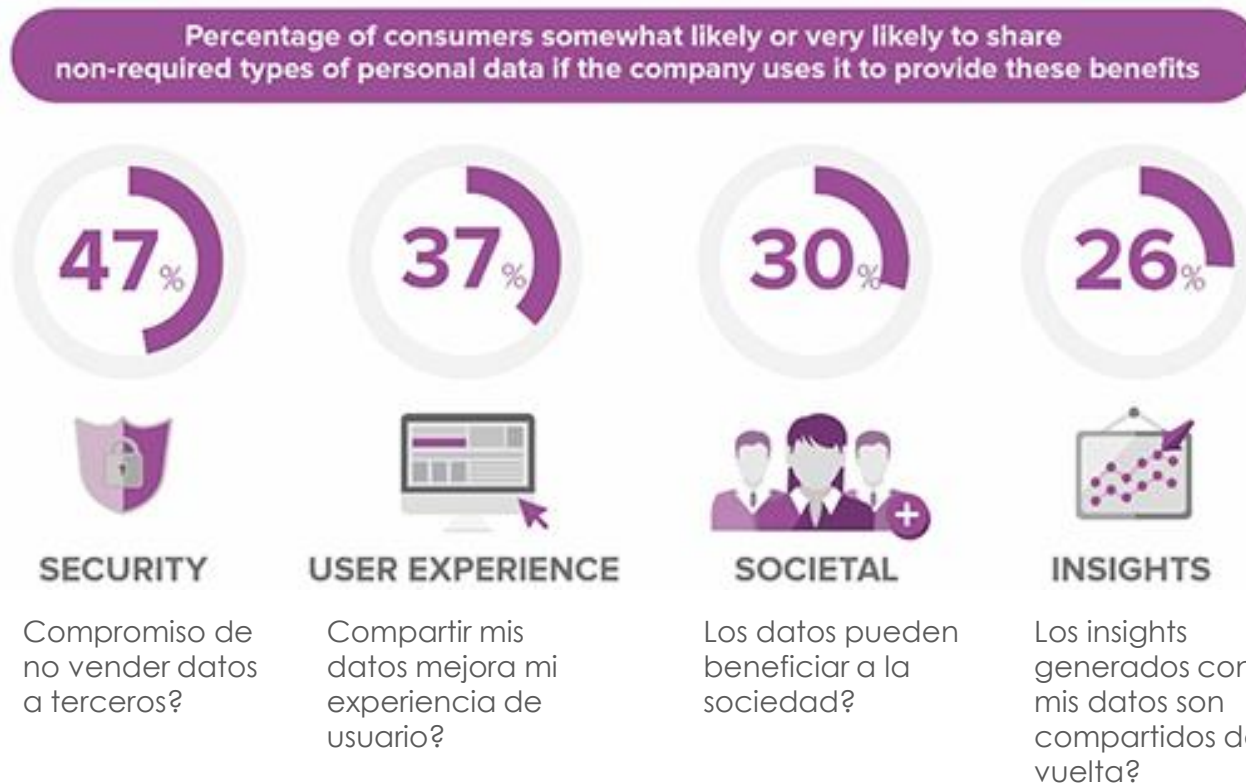
HAPPY GO LUCKY

Low Defense, Happy To Share:

Millennials	10%
Gen X	11%
Boomers	10%
Silent Gen	13%

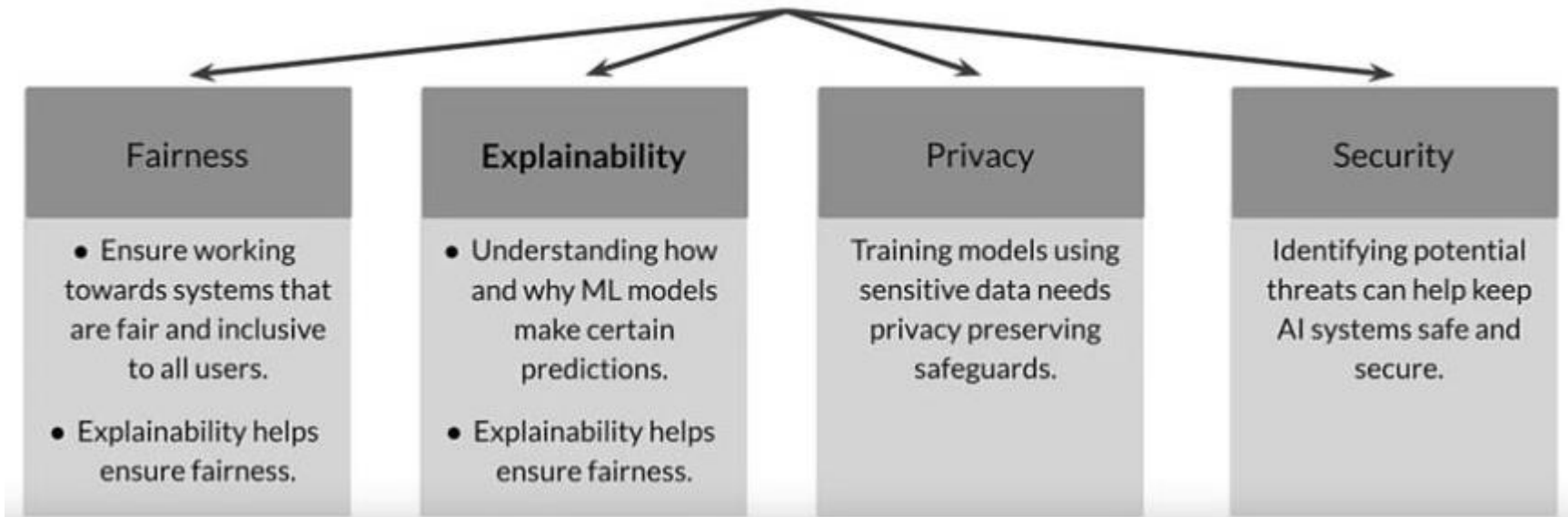
Principales retos

Que ofrecer para que los clientes compartan su información?



IA Responsable

- El uso del IA crea nuevas oportunidades de mejorar negocios y la vida de las personas.
- Pero también levanta nuevas preguntas sobre la mejor forma de construir estos nuevos sistemas IA.





Clase 04: Aplicaciones en la Industria y Fuga de Clientes

FUGA DE CLIENTES

Churn

- Término utilizado para referirse a la **deserción** o a la **pérdida** de un cliente.
- Clientes o contratos que **cortan** el **vínculo** con un servicio o una empresa durante un período de **tiempo dado**.



Tasa de Abandono

- La **tasa de abandono** puede ser representada de distintas maneras:



Número de
clientes
perdidos



Valor de los
negocios
recurrentes
perdidos



% de clientes
perdidos

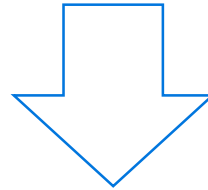


% de valor
recurrente
perdido



Prevención

- Adquisición de clientes **más leales** desde un inicio
- Identificación de los clientes con **mayores posibilidades** de desertar



Modelo predictivo de
fuga

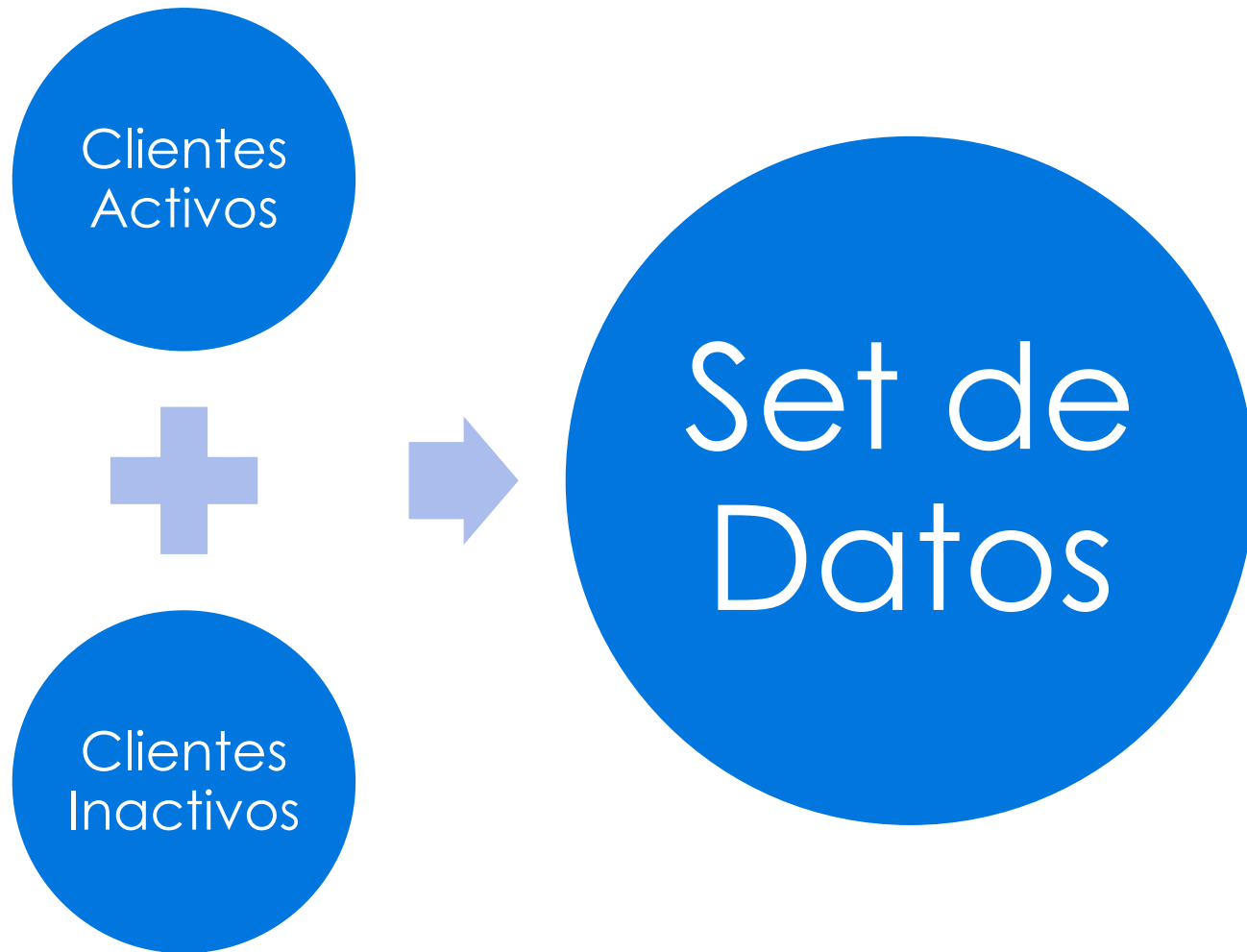
Definir actividad y abandono

- Es necesario **definir** claramente qué es un **cliente activo** y qué es un **cliente fugado** (o fugando)



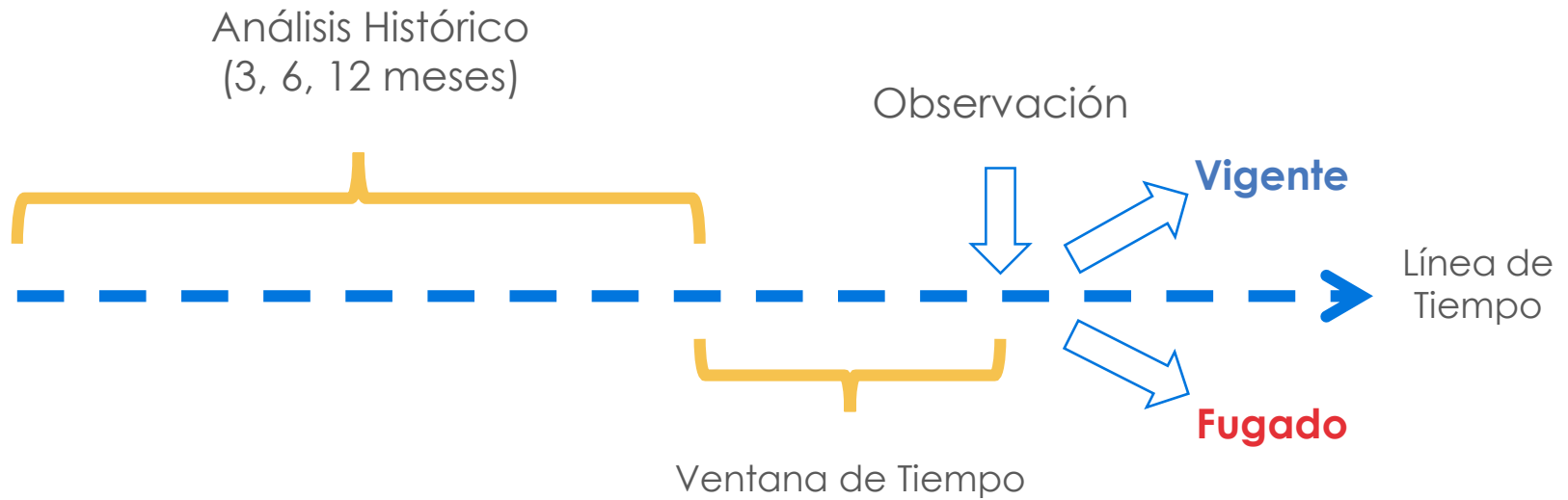


Consolidar los datos



Establecer marco temporal de predicción

- ¿Con cuánto tiempo de **anticipación** se debe hacer la predicción?
- Nota: Es necesario establecer cuánto es el tiempo que se requiere para tomar **acciones comerciales**



Luego...

- Aplicar algunas de las metodologías de análisis de datos ya aprendidas





Clase 04: Aplicaciones en la Industria y Fuga de Clientes

MARKETING



Marketing

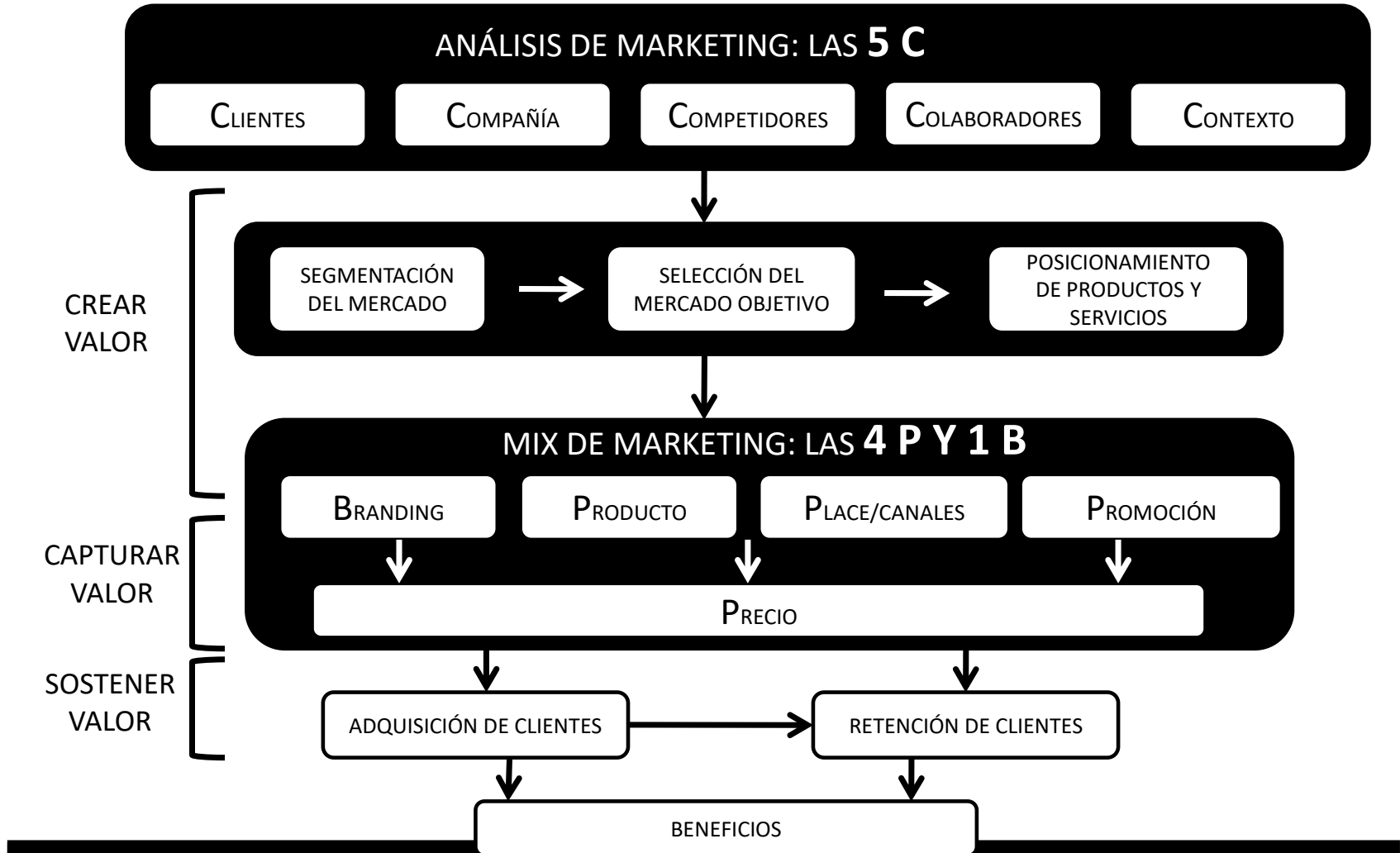
- Proceso por el cual una empresa **crea valor** para sus clientes elegidos. Una vez creado el valor para sus clientes, la empresa está autorizada a **capturar una porción** del mismo a través del precio.
- Asimismo, el marketing se debe ocupar de la **sostenibilidad de la propuesta de valor** para el cliente, para asegurar la pervivencia de la empresa en el tiempo.



“Una empresa sólo tiene dos funciones básicas: Marketing e Innovación”

Peter Drucker

Modelo estrategia de marketing





Analizar y escuchar a clientes

- **Conocer y segmentar** a los clientes tomando como base distintos parámetros. **Personalizar la oferta** como **aumentar el engagement** con el cliente a través de los datos históricos almacenados.
- Implementación de **soluciones de escucha y fidelización** de clientes (encuestas de satisfacción, regalos promocionales, etc.) pueden suponer el impulso necesario a la decisión del consumidor de seguir confiando en nosotros.



Clase 04: Aplicaciones en la Industria y Fuga de Clientes

PRIMER PROYECTO



Descripción del Proyecto

- Construir modelo predictivo de fuga de clientes en industria de telecomunicaciones.
- Preparación de datos, contexto, análisis descriptivo, nuevas variables.
- Recorrer el espacio de parámetros del algoritmo, robustez y presentación de resultados.
- Proponer 4 acciones para retener a los clientes

Como mínimo: Análisis exploratorio, limpieza de datos, ingeniería de features, entrenamiento de modelo usando train dataset y métodos árbol de decisión, SVM y/o random forest, ajuste de parámetros, validación del modelo usando datos de test dataset, presentación e interpretación de resultados, propuestas retención.



Evaluación

- Criterios de evaluación

Aspectos a ser evaluados	Ponderación en la Nota Final
Presentación: Estructura, claridad de contenido y atractivo visual de la ppt	30%
Presentación: Descripción del problema, metodología aplicada, resultados y propuesta de acciones de retención	70%

- Entrega

Las presentaciones deben ser enviadas hasta las 23:59 horas del día martes 4 de abril a los correos mecastillo2@uc.cl y regonzar@uc.cl

- Máximo 15 slides

Las presentaciones se realizarán durante la clase del 5 de Abril.

- 15 minutos presentación + 3 minutos preguntas



ESCUELA DE INGENIERÍA
FACULTAD DE INGENIERÍA

EDUCACIÓN
PROFESIONAL

Diplomado en Big Data y Ciencia de Datos
Ciencia de Datos y sus Aplicaciones

Clase 04:

Aplicaciones en la Industria y Fuga de Clientes

Roberto González



regonzar@uc.cl

