



ESCUELA DE INGENIERÍA  
FACULTAD DE INGENIERÍA

EDUCACIÓN  
PROFESIONAL

# Diplomado en Big Data y Ciencia de Datos

## *Curso: Ciencia de Datos y sus Aplicaciones*

### Clase 04: Aplicaciones en la Industria y Fuga de Clientes

Jaime Caiceo, Roberto González, Roberto Muñoz

✉ regonzar@uc.cl

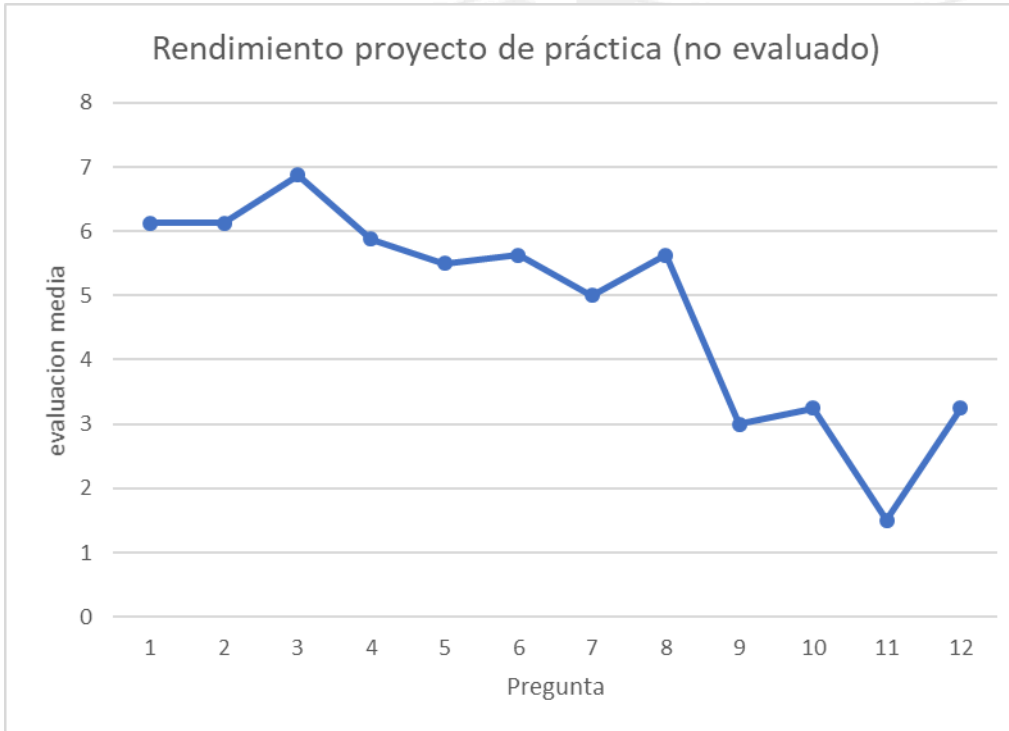
✉ jcaiceo@uc.cl

✉ rmunoz@uc.cl

El uso de apuntes de clases estará reservado para finalidades académicas. La reproducción total o parcial de los mismos por cualquier medio, así como su difusión y distribución a terceras personas no está permitida, salvo con autorización del autor.



## Feedback trabajo anterior (valor dolar)



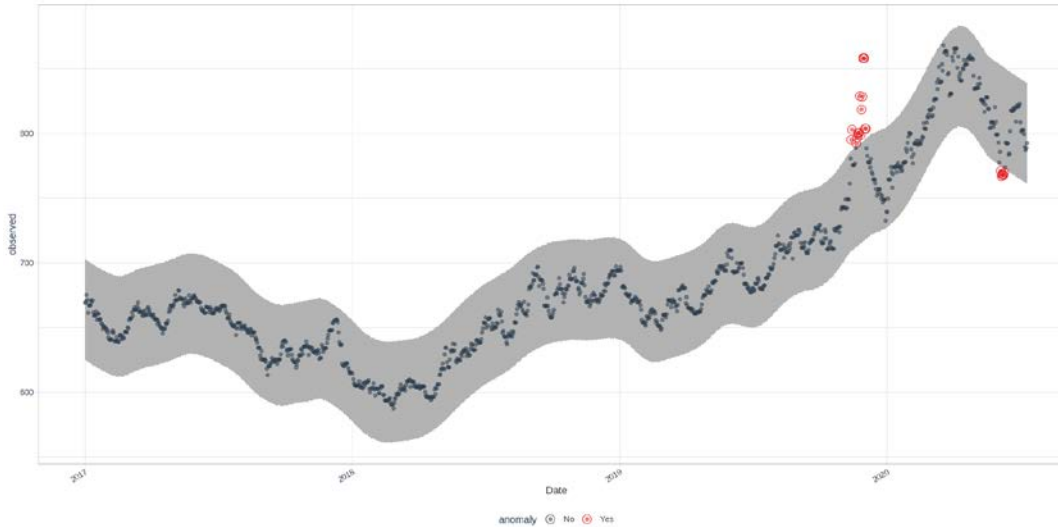
Top 3 grupos:

3 → felicidades(respondieron todo)

4

8

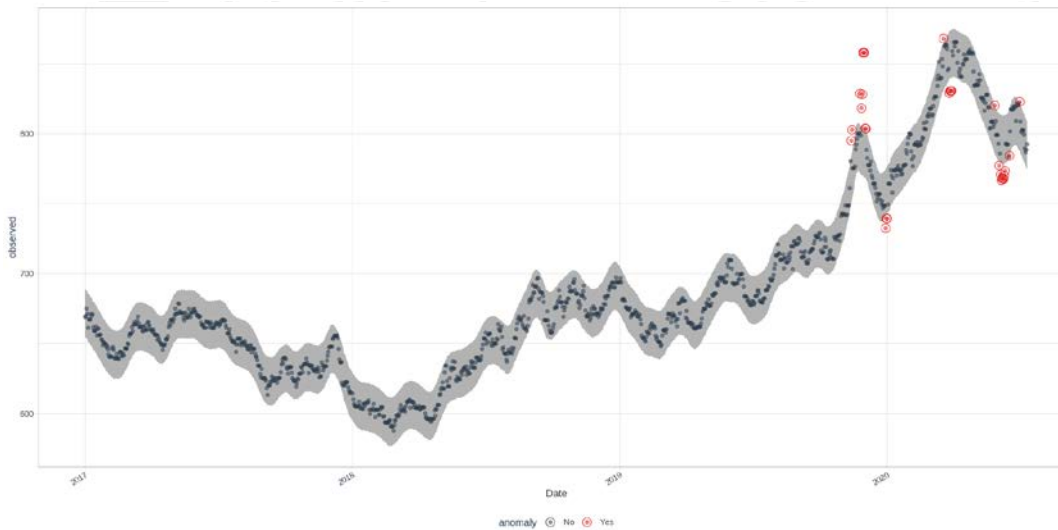
1. Explique el metodo usado en esta actividad para determinar los outliers del valor del dolar. ¿Cuál sería el contexto en este caso?
2. ¿Cuales son las ventajas y ventajas de este metodo en comparacion al analisis puntual y contextual basado en el uso de cuartiles del boxplot?
3. Modifique los parametros `frequency` y `trend` de la funcion `time_decompose()` y determine los valores de estos parametros que entregan los mejores resultados. ¿Cual es el criterio que usa para definir que estos resultados son mejores?
4. ¿Considera que los resultados de este metodo son suficientemente buenos para implementarlos como una solución productiva en la empresa? ¿Por qué?



### Actividad 3

frequency = 7 days

trend = 91 days



trend = 30 days



ESCUELA DE INGENIERÍA  
FACULTAD DE INGENIERÍA

EDUCACIÓN  
PROFESIONAL

Clase 04: Aplicaciones en la Industria y Fuga de Clientes

# INDUSTRIA GUIADA POR DATOS

# Data Scientist

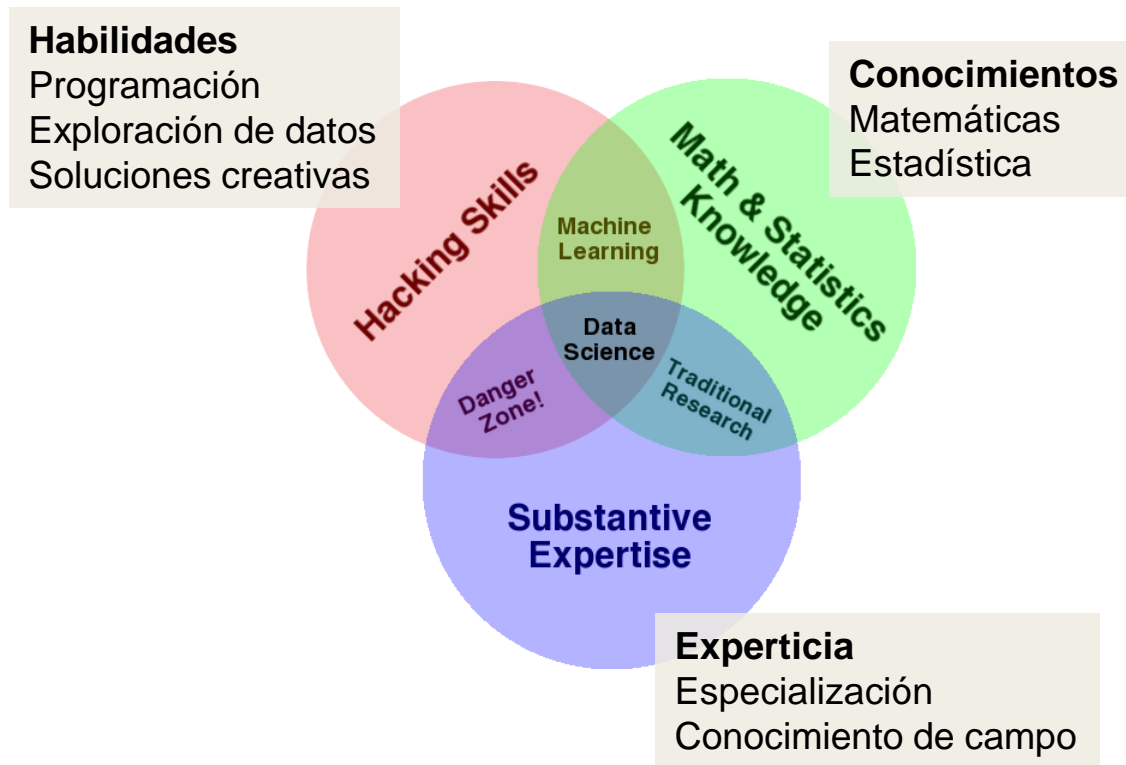
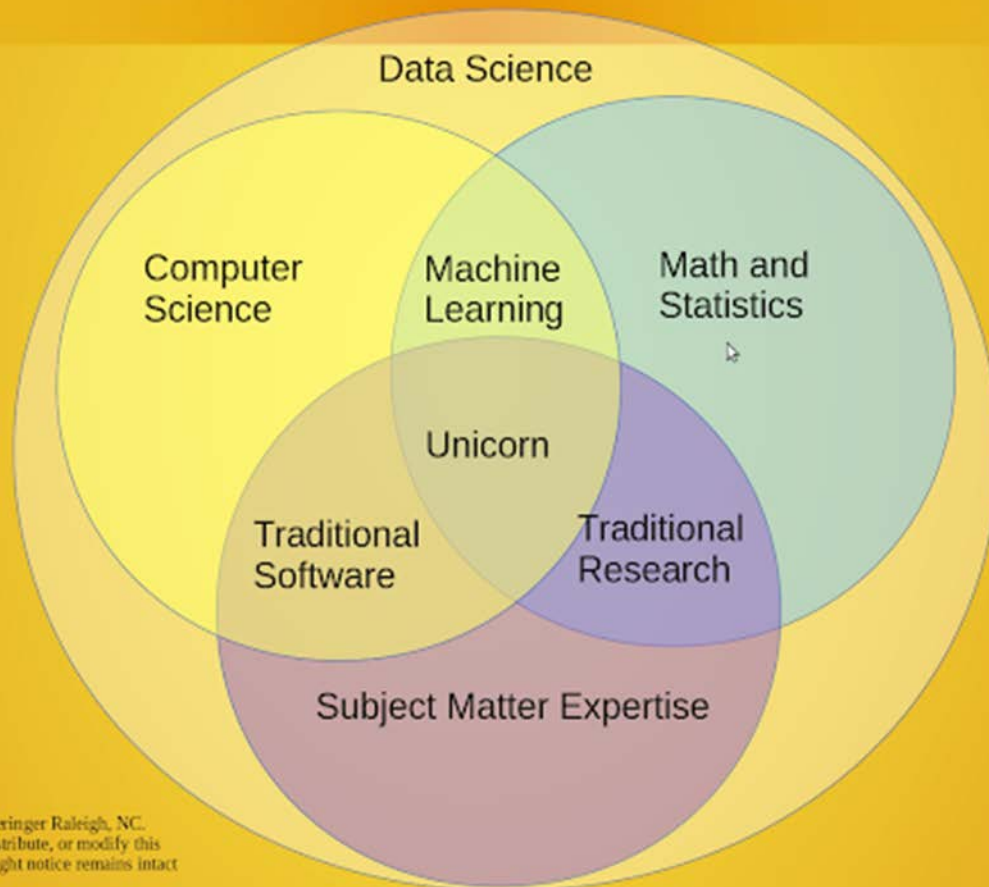


Diagrama de Venn para Data Science  
Drew Conway (2010)



# 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

# 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.







ESCUELA DE INGENIERÍA  
FACULTAD DE INGENIERÍA

EDUCACIÓN  
PROFESIONAL

Clase 04: Aplicaciones en la Industria y Fuga de Clientes

# APLICACIONES



ESCUELA DE INGENIERÍA  
FACULTAD DE INGENIERÍA

EDUCACIÓN  
PROFESIONAL

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



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

## Canales



### Voz

Voice to Text  
Speech recognition  
NLP/NLG



### Texto & Email

Text Mining  
Data Mining  
Semantic analysis / Sentiment Analysis



### Documentos & Medios

OCR  
Reconocimiento de imágenes y videos



Advanced  
Analytics



Data  
Management



RPA



Data  
Visualization

## Resultados



### Customer Interaction

Channel Management



Chatbots & Virtual Assistant



### Analytics

Deep learning  
Reinforcement learning  
Unsupervised learning  
Supervised learning  
NLP & Sentiment Analysis



### Decisioning

ML & Decision Analytics  
Prescriptive Analytics



Knowledge graphs  
Recommender systems





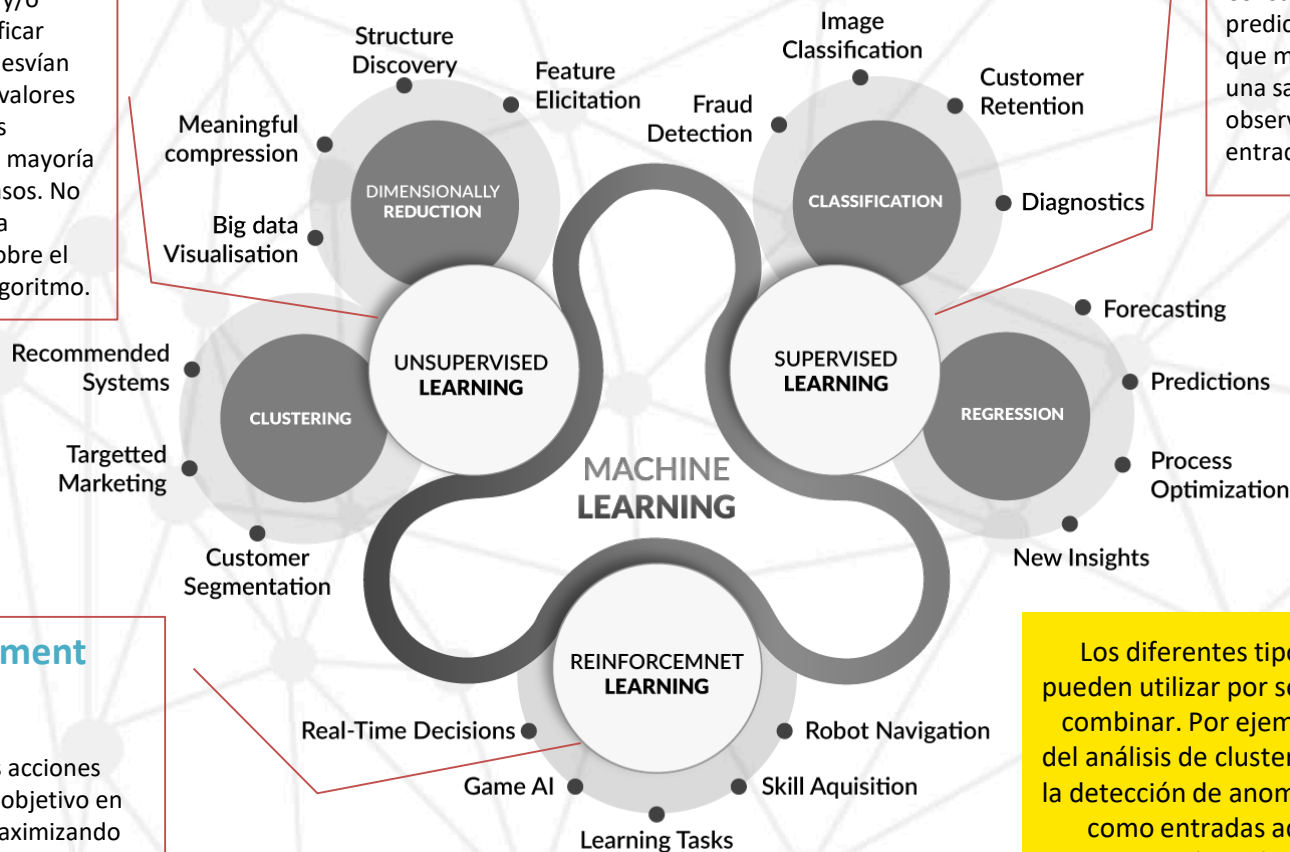
## Enfoque para los distintos tipos de aprendizaje

### 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



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

## Telecomunicaciones







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

## Energy & Utilities





## Casos de Uso de Machine Learning

Computación basada en reglas		Machine Learning			Inteligencia de Máquina General
Modo Cognitivo	Inferencia basada en reglas	Supervised Learning	Unsupervised Narrow Learning	Unsupervised Context Aware Learning	Self Aware Unsupervised Learning
Natural Language Processing 	Comprobación de ortografía y gramática	Dictado de voz a texto	Asistente personal para Q&A básicas basadas en voz	Diálogo y traducción en tiempo real	Idioma, sarcasmo, matices, entonación
Computer Vision 	Inspección de defectos en la fruta con imágenes infrarrojas	Reconocimiento facial	Clasificación compleja (Ej: búsqueda de segmentos de video)	Sistemas de visión para vehículos autónomos	Agentes de seguridad digital Agentes de exploración autónomos
Pattern Recognition 	Inspección industrial basada en reglas por funcionamiento defectuoso	Detección de fraude (Basado en patrones históricos)	Recomendación de productos basada en la preferencia del cliente	Diagnóstico clínico automatizado en tiempo real	Predicción de la evolución de las enfermedades
Reasoning & Optimization 	Mantenimiento de diagnóstico	Mantenimiento predictivo para maquinaria	Predicción de fallos en los sistemas críticos	Automatizar las recomendaciones sobre los insumos en la cadena de valor	Motor de búsqueda capaz de responder preguntas en lugar de sólo presentar resultados
Más de 5 años atrás		5 años atrás		Hoy	Futuro



# 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"><li>▶ Mejora la interacción con el cliente</li><li>▶ Simplificación de tareas</li><li>▶ Personalización del servicio</li></ul>	<ul style="list-style-type: none"><li>▶ Recomendaciones personalizadas</li><li>▶ Gestión de leads</li><li>▶ Ofertas a la medida</li><li>▶ Relación entre producto y cliente</li></ul>	<ul style="list-style-type: none"><li>▶ Fraudes</li><li>▶ Garantizar la fiabilidad y la calidad del servicio</li><li>▶ Autenticación e identidad</li><li>▶ Sistemas de reputación</li><li>▶ Colocación de créditos</li></ul>	<ul style="list-style-type: none"><li>▶ Gestión de las finanzas personales</li><li>▶ Análisis de riesgos</li><li>▶ Perspectivas del mercado y la competencia</li><li>▶ Análisis de tendencias</li></ul>	<ul style="list-style-type: none"><li>▶ Operaciones automatizadas avanzadas</li><li>▶ Seguridad mejorada en un entorno abierto</li></ul>



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

<https://pollev.com/robertoeugen941>

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

### The Data-Sharing Mindsets of Consumers

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





ESCUELA DE INGENIERÍA  
FACULTAD DE INGENIERÍA

EDUCACIÓN  
PROFESIONAL

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

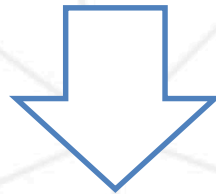


**% de valor**  
recurrente perdido



## 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

Cientes  
Activos



Cientes  
Inactivos

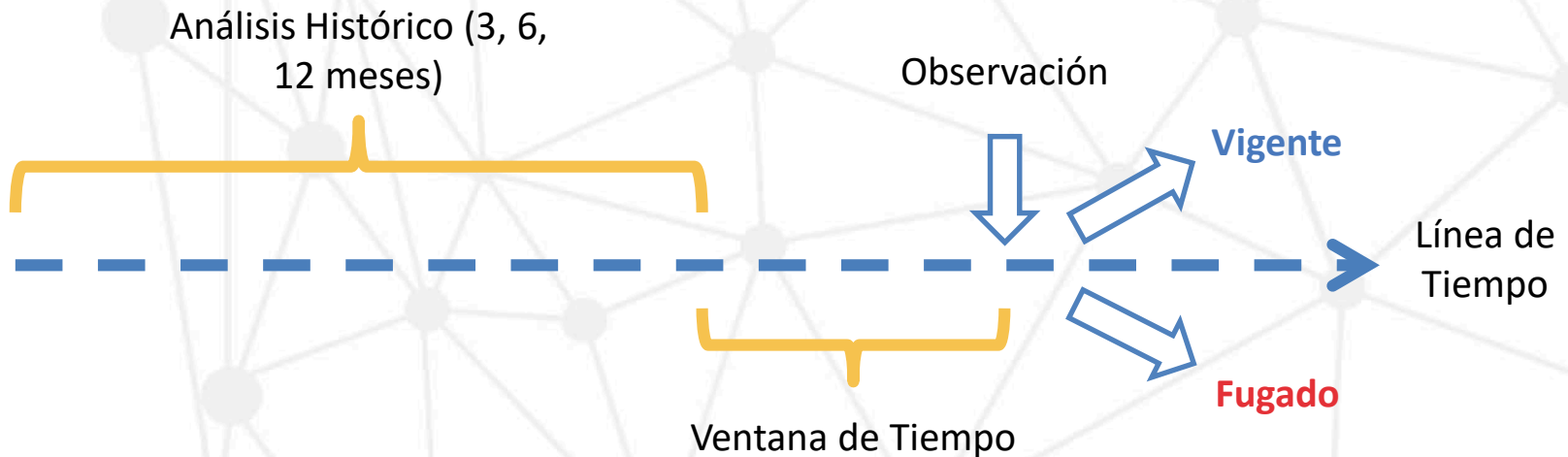
Set de  
Datos





## 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





ESCUELA DE INGENIERÍA  
FACULTAD DE INGENIERÍA

EDUCACIÓN  
PROFESIONAL

Clase 04: Aplicaciones en la Industria y Fuga de Clientes

# PRIMER PROYECTO

Diplomado en Big Data y Ciencia de Datos / © Roberto  
González, Roberto Muñoz, Jaime Caiceo / 2021



## Descripción del Proyecto

- Construir modelo predictivo de fuga de clientes en industria de telecomunicaciones.
- Recorrer el espacio de parámetros del algoritmo y anotar resultados.
- Proponer 4 acciones para retener a los clientes

**Sesión 1:** Análisis exploratorio, limpieza de datos, ingeniería de features

**Sesión 2:** Entrenamiento de modelo usando train dataset y métodos árbol de decisión, SVM o random forest. Validación del modelo usando datos de test dataset.



## Evaluación

- Criterios de evaluación

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

- Entrega

Las presentaciones deben ser enviadas hasta las 23:59 horas del martes 13 de abril a los correos [regonzar@uc.cl](mailto:regonzar@uc.cl) , [rmunoz@uc.cl](mailto:rmunoz@uc.cl) y [icaiceo@uc.cl](mailto:icaiceo@uc.cl)

Las presentaciones se realizarán durante la clase del 14 Abril





ESCUELA DE INGENIERÍA  
FACULTAD DE INGENIERÍA

EDUCACIÓN  
PROFESIONAL

# Diplomado en Big Data y Ciencia de Datos

## *Curso: Ciencia de Datos y sus Aplicaciones*

### Clase 04: Aplicaciones en la Industria y Fuga de Clientes

Jaime Caiceo, Roberto González, Roberto Muñoz

✉ regonzar@uc.cl

✉ jcaiceo@uc.cl

✉ rmunoz@uc.cl

El uso de apuntes de clases estará reservado para finalidades académicas. La reproducción total o parcial de los mismos por cualquier medio, así como su difusión y distribución a terceras personas no está permitida, salvo con autorización del autor.