



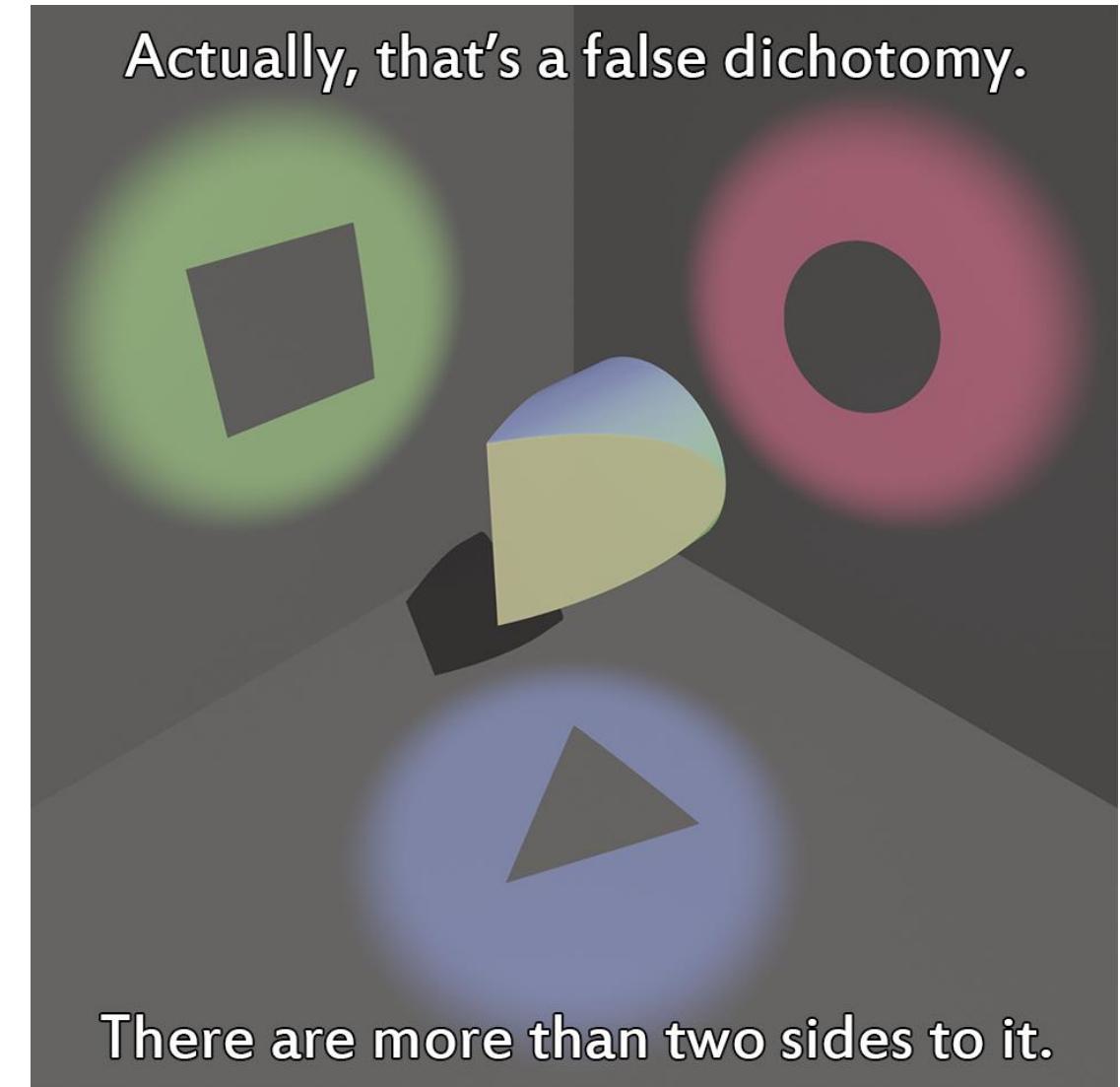
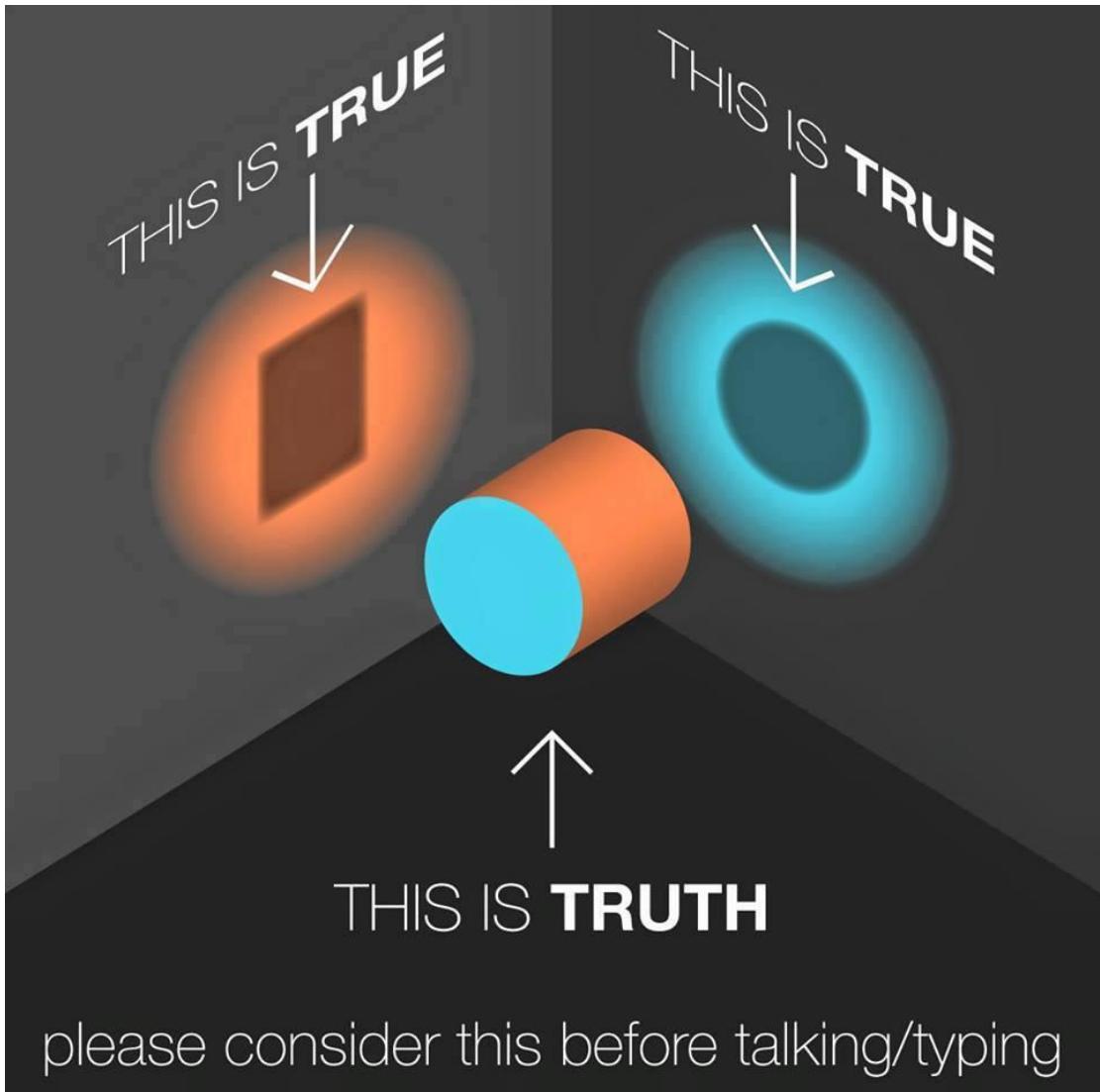
teradata.

Decisiones Data Driven, ¿Qué nos falta?

Luis Cajachahua
Senior Data Scientist
Center of Excellence, Teradata Americas

Junio 2020

Antes de empezar...



Gartner predijo que:
“El 60% de los proyectos de big data hasta 2017, no irán más allá de la experimentación y serán abandonados”

Gartner.

Newsroom

Press Release

Share: [Tweet](#) [Share](#) 471 [G+](#)

Egham, UK, September 15, 2015

[View All Press Releases](#)

Gartner Says Business Intelligence and Analytics Leaders Must Focus on Mindsets and Culture to Kick Start Advanced Analytics

Analysts to Explore Advanced Analytics at the Gartner Business Intelligence & Analytics Summit 2015, October 14-15 in Munich, Germany

Getting started with [advanced analytics](#) is as much about changing mindsets and culture as it is about acquiring tools and skills, according to Gartner, Inc. Failure to make these changes can be fatal to success. Gartner predicts that, through 2017, 60 percent of [big data](#) projects will fail to go beyond piloting and experimentation, and will be abandoned.

<http://www.gartner.com/newsroom/id/3130017>



Nick Heudecker

@nheudecker

Seguir



We were too conservative. The failure rate is closer to 85%. And the problem isn't technology.

Ronda Swaney @RondaSwaney

Gartner says in 2017, 60% of #bigdata projects will fail to move past preliminary stages. Does that mean big data has been a failure? No. But it does mean we have to separate hype from reality. Read my latest written for @Hortonworks.
hortonworks.com/article/what-h... via @Hortonworks

Traducir Tweet

12:25 - 9 nov. 2017

26 Retweets 22 Me gusta



4

26

22



“Fuimos demasiado conservadores. La tasa de fallos está cerca del 85%. Y el problema NO es la tecnología.”

(Nick Heudecker, Analista de Gartner)

teradata.

<https://designingforanalytics.com/resources/failure-rates-for-analytics-bi-iot-and-big-data-projects-85-yikes/>

Our Top Data and Analytics Predicts for 2019

by **Andrew White** | January 3, 2019 | Comments Off on Our Top Data and Analytics Predicts for 2019

Predicts 2019: Data and Analytics Strategy

- By 2022, 90% of corporate strategies will explicitly mention information as a critical enterprise asset and analytics as an essential competency.
 - By 2023, data literacy will become an explicit and necessary driver of business value, demonstrated by its formal inclusion in over 80% of data and analytics strategies and change management programs.
 - By 2022, 30% of CDOs will partner with their CFO to formally value the organization's information assets for improved information management and benefits.
 - By 2023, 60% of organizations with more than 20 data scientists will require a professional code of conduct incorporating ethical use of data and AI.
 - By 2022, more than half of major new business systems will incorporate continuous intelligence that uses real-time context data to improve decisions.

Predicts 2019: Analytics and BI Solutions

- Through 2020, 80% of AI projects will remain alchemy, run by wizards whose talents will not scale in the organization.
 - Through 2022, only 20% of analytic insights will deliver business outcomes.
 - By 2021, proof-of-concept analytic projects using quantum computing infrastructure will have outperformed traditional analytic approaches in multiple domains by at least a factor of 10

Para el 2023, la alfabetización de datos se convertirá en un impulsor explícito y necesario del valor comercial

Código de conducta para el uso ético de los Datos y de la IA

Hasta 2020, el 80% de los proyectos de IA seguirán siendo alquimia, a cargo de magos, cuyos talentos no escalarán en la organización.

Hasta 2022, solo el 20% de los insights analíticos generarán resultados comerciales.

Strategic Planning Assumptions

By 2022, augmented analytics technology will be ubiquitous, but only 10% of analysts will use its full potential.

By 2022, 40% of machine learning model development and scoring will be done in products that do not have machine learning as their primary goal.

By 2023, 90% of the world's top 500 companies will have converged analytics governance into broader data and analytics governance initiatives.

By 2025, 80% of consumer or industrial products containing electronics will perform analysis on the device.

The New Equations for Analytic Success

$$\text{Untrained Human} + \text{Smart Analytics} = \text{Untrained Human} + \text{Unused Analytics}$$

Augmented analytics will be ubiquitous, but few users will tap its full potential.

$$\text{Anything} + \text{Machine Learning} = \text{Everything} + \text{Machine Learning}$$

Machine learning will increasingly be found in products with differing primary goals.

$$\text{Data Governance} + \text{Analytics Governance} = (\text{Data and Analytics}) \text{ Governance}$$

Companies will converge analytics governance and data and analytics governance.

$$\text{Dumb Device} + \text{Analytics} + \text{New Skills} = \text{Smart Device}$$

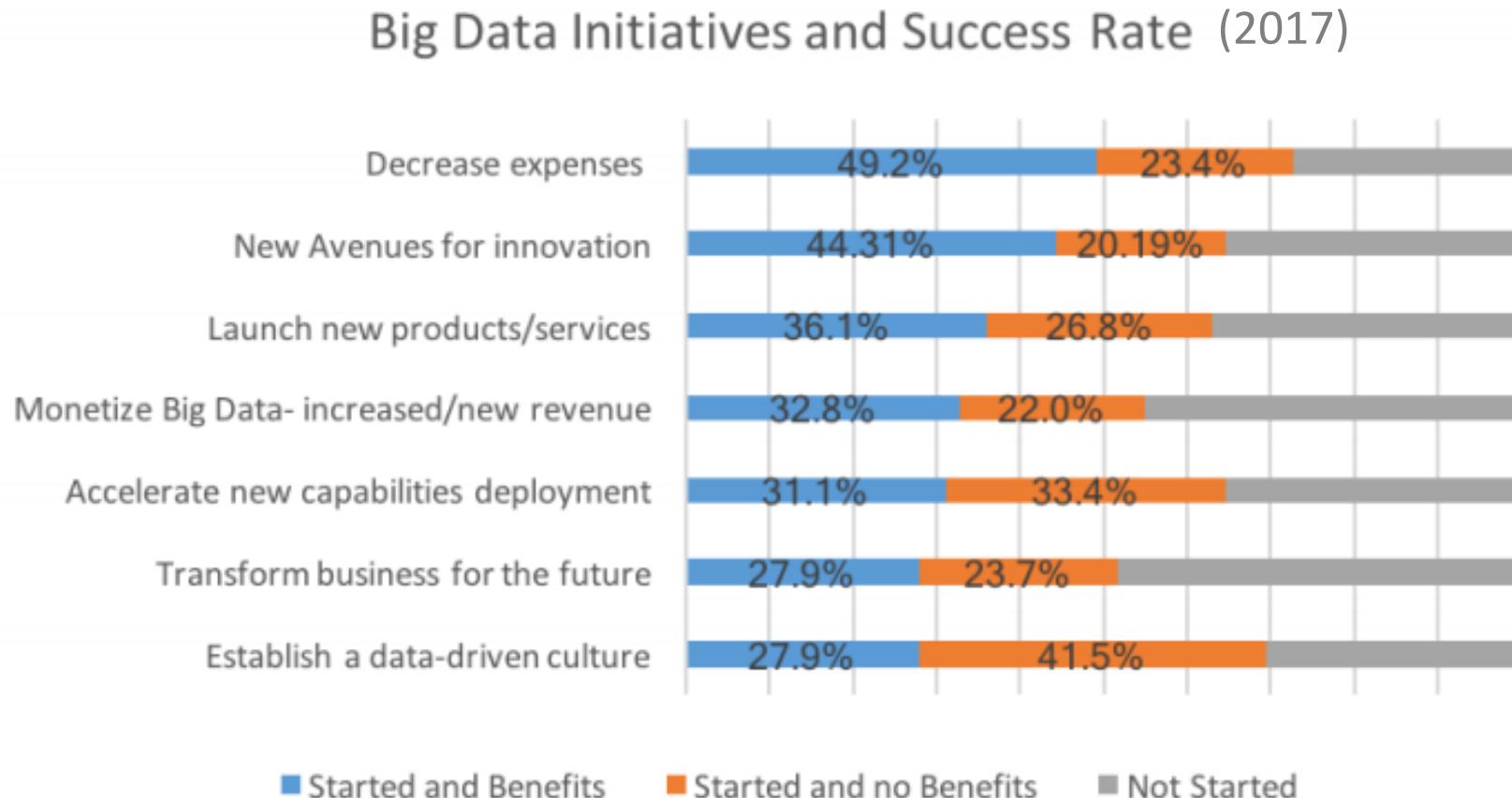
Most products containing electronics will incorporate on-device analytics.

Sin usuarios entrenados, Smart Analytics no se utilizará en todo su potencial

Los esfuerzos de Gobierno de Datos y Analytics convergerán

Muchos productos incluirán Analytics en el dispositivo

Los ejecutivos responden (NVP Survey 2017)



69.4% de las empresas encuestadas tuvieron iniciativas para implementar una cultura Data Driven, pero sólo el **27.9%** tuvo éxito.

Los ejecutivos responden (NVP Survey 2019)

<u>The State of Big Data/AI in 2019</u>	<u>Yes</u>	<u>No</u>
Managing Data as a Business Asset	46.9%	53.1%
Forged a Data Culture	28.3%	71.7%
Driving Innovation with Data	59.5%	40.5%
Competing on Data and Analytics	47.6%	52.4%
Created a Data-Driven Organization	31.0%	69.0%

On every metric except driving innovation with data, firms ranked themselves as failing to transform their businesses:

- 71.7% of firms report that they have yet to forge a data culture
- 69.0% of firms report that they have not created a data-driven culture
- 53.1% of firms state they are not yet treating data as a business asset
- 52.4% of firms claim that they are not competing on data and analytics.

El 69% de las empresas informan que no han creado una cultura data-driven.

<u>Principle Challenge to Becoming Data-Driven</u>	<u>2018</u>	<u>2019</u>
People	48.5%	62.5%
Process	32.4%	30.0%
Technology	19.1%	7.5%

Obstáculos para convertirse en una organización Data Driven: 92.5% opinan que tiene que ver con Personas y Procesos

It is notable that organizations identifying cultural challenges – people and process – continues to accelerate, representing 92.5% of the gap this year.

¿Cuáles son las barreras, entonces?

- **Tecnología:**
 - Superabundancia de herramientas y tecnologías que nacen y son reemplazadas en corto tiempo. ¿Soluciones o Buzzwords?
 - Negocio Open Source: Si no pagas por el automóvil terminado, necesitarás muchas piezas, personas que lo armen y mantengan.
 - Negocio nube: +componentes=+consumo.
 - Retos en el despliegue en producción.
- **Profesionales:**
 - Escasez de talento (Cacería de Unicornios).
 - Dificultad para acceder a entrenamiento de calidad, con buenas prácticas.
 - ¿Es más importante utilizar la tecnología de moda o solucionar el problema?
- **Organizaciones:**
 - Presiones para reducir costos.
 - Falta de Liderazgo y líos políticos.
 - Objetivos poco realistas y contradictorios.
 - ¿Realmente entienden a los clientes?
 - Modernización de tecnologías sin cambio de procesos ni cultura (CRM, Digital, CX, etc.)
 - Limitaciones para retener el talento, generar líneas de carrera y desarrollar competencias.
- **Datos:**
 - Silos de información (Pobre integración)
 - Cantidad pero no Calidad.
 - Dificultad para obtenerlos oportunamente.
 - Los datos que faltan y nunca se recolectan.

¿Y qué puedo hacer yo?

Ser parte de la solución

Recordemos: Todo tiene que ver con las personas

- Competencias
 - Talento
 - Formación (Saber)
 - Entrenamiento y práctica (Saber hacer)
 - Saber vender (¡Todos vendemos!)
 - Buenas prácticas (Aprender de otros)
 - Trabajo en Equipo
 - Espíritu crítico
 - Resiliencia (Fail fast)
 - Pasión



Pero lo más importante ¿Qué quieres tú?



C-Level/Manager con
interés en Analytics



Data Engineer



Profesional en IT/SW



Analista de Negocio



Arquitecto



Estudiante en
Formación



Profesional de otro rubro
trabajando en un equipo
Analítico



Analista de Datos



Recomendaciones Generales

Cómo convertirse en Data Scientist Hacker

← → C sindominio.net/biblioweb/telematica/hacker-como.html ☆ ⚡ O 🔍 📁 | 🌐

La actitud del hacker

Los hackers resuelven problemas y construyen cosas, y creen en la libertad y la ayuda voluntaria mutua. Para ser aceptado como hacker, deberás comportarte como si tuvieras esta actitud en tu interior. Y para comportarte como si tuvieras esta actitud, deberás creerte de verdad dicha actitud.

Pero si piensas en cultivar las actitudes de hacker solo como una forma de ganar aceptación en esta cultura, te estás equivocando. Transformarse en la clase de persona que cree estas cosas es importante para *ti* —para ayudarte a aprender y mantenerte motivado. Como en todas las artes creativas, el modo más efectivo de transformarse en un maestro es imitar la mentalidad de los maestros —no sólo intelectualmente, sino también emocionalmente.

O como dice el siguiente poema zen moderno:

Para seguir la trayectoria:
mira al maestro,
sigue al maestro,
camina junto con el maestro,
mira a través del maestro,
conviértete en el maestro.

Así que, si quieras ser un hacker, repite lo que sigue hasta que te creas lo que estás diciendo:

Para seguir la trayectoria:
mira al maestro,
sigue al maestro,
camina junto con el maestro,
mira a través del maestro,
conviértete en el maestro.

Recomendaciones Generales

- Elige fuentes confiables. Buscar de acuerdo al perfil y al nivel en el que te encuentras.
- El mejor material siempre está en inglés.
- Lo más popular no necesariamente es lo mejor.



Conocimiento Especializado

Skills Técnicos

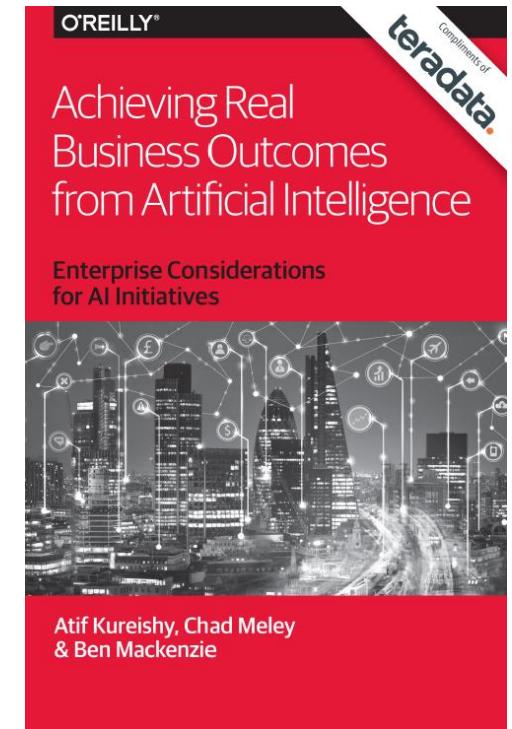
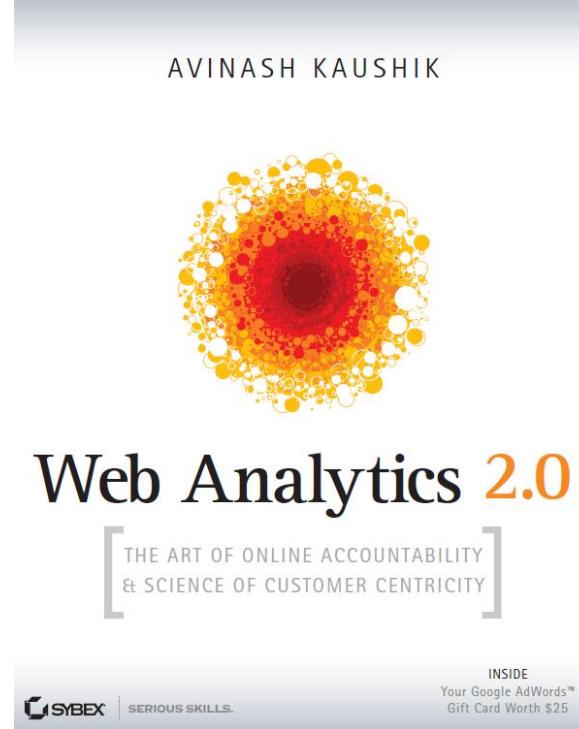
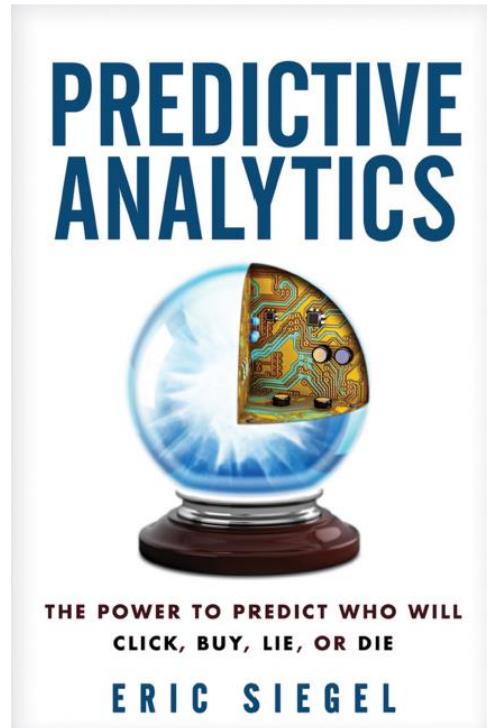
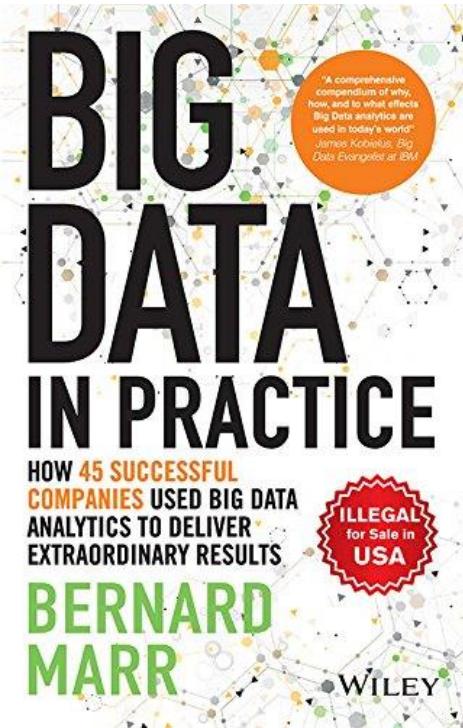
Matemática y Estadística

Soft Skills

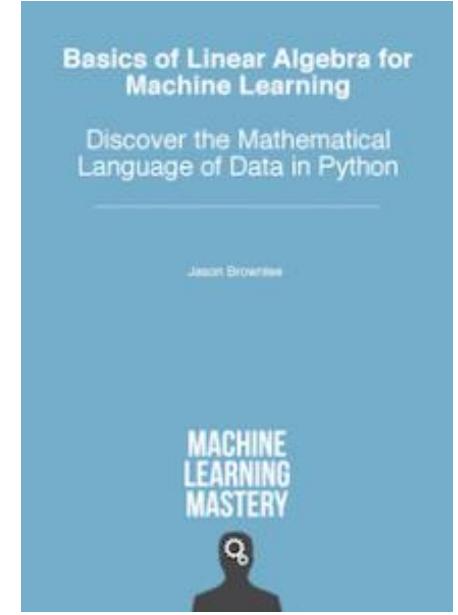
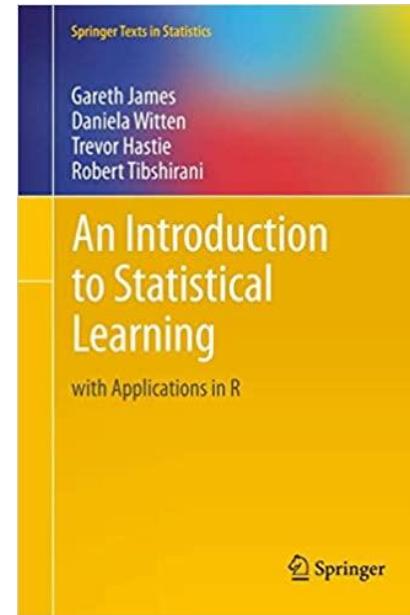
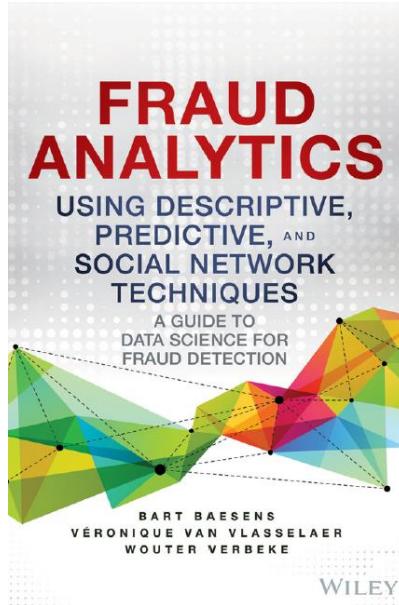
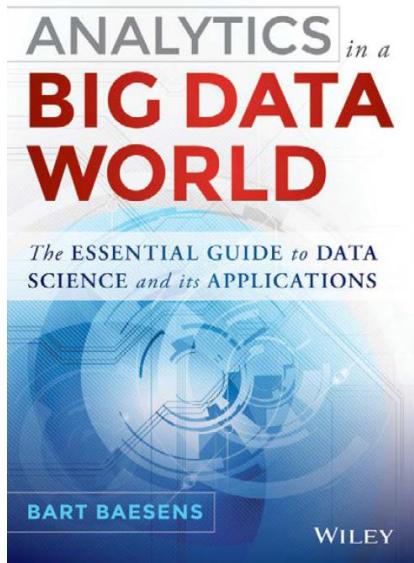
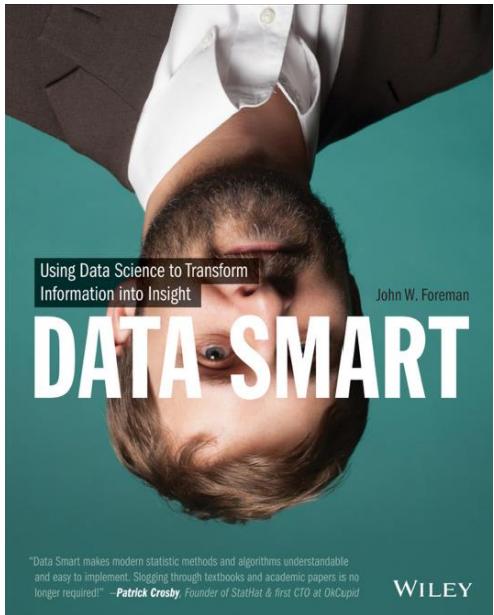
Pensamiento Analítico

Sentido Común

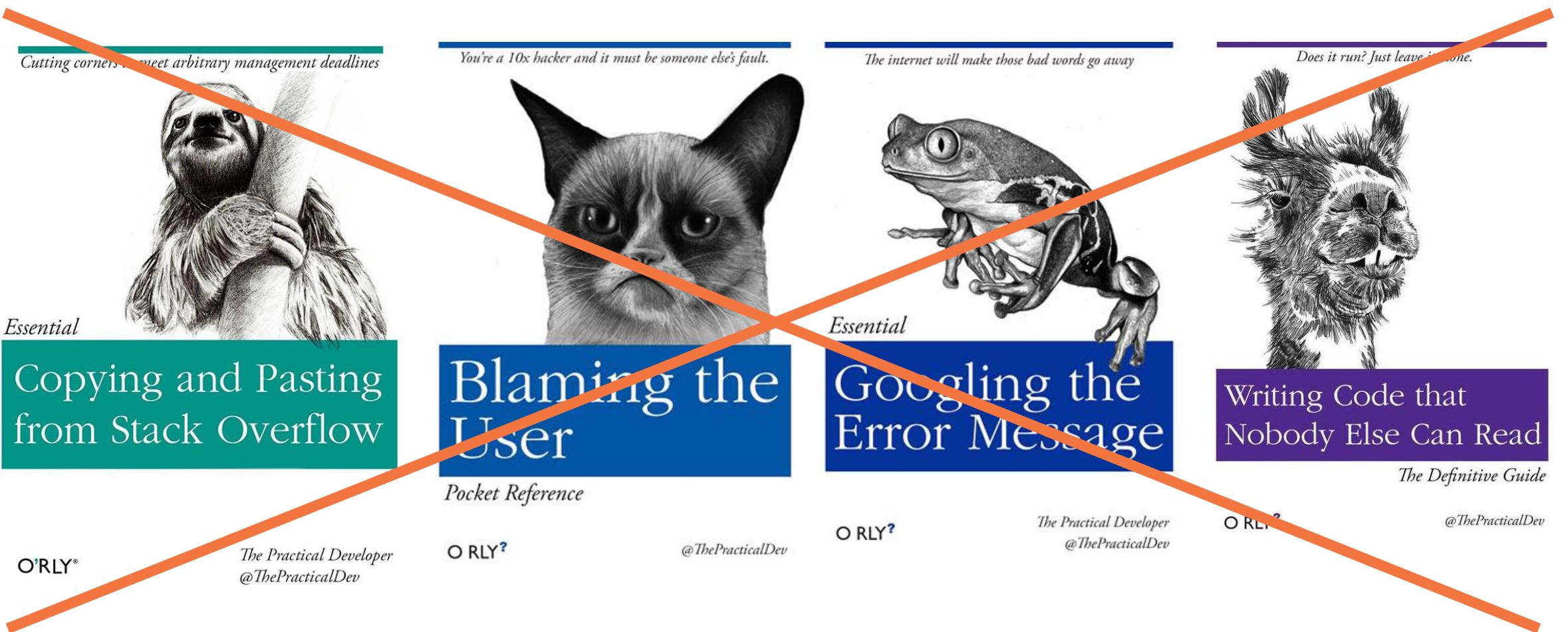
Colección para Managers



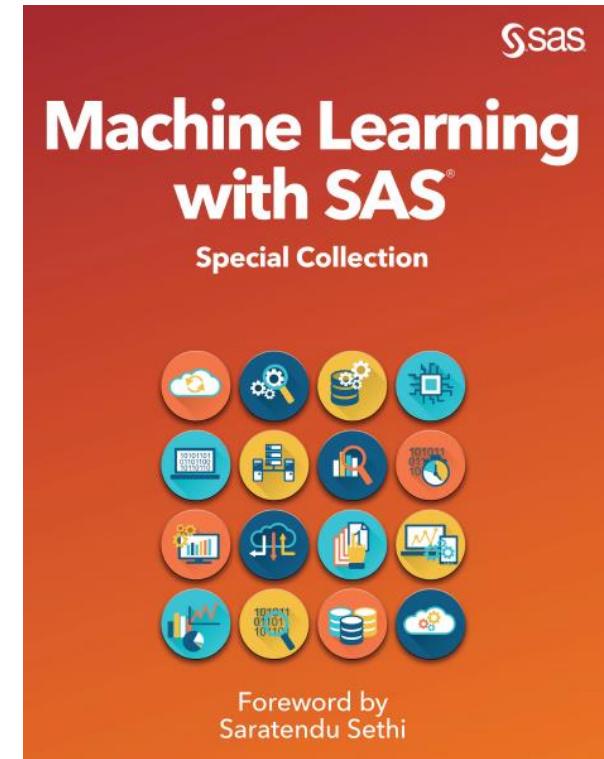
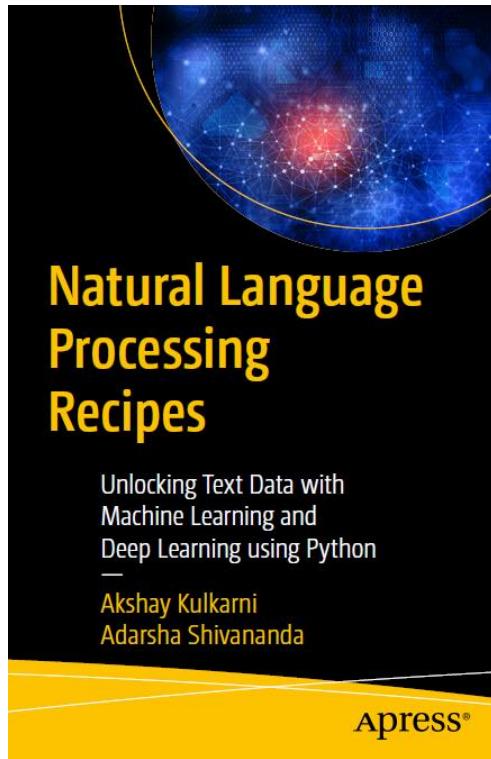
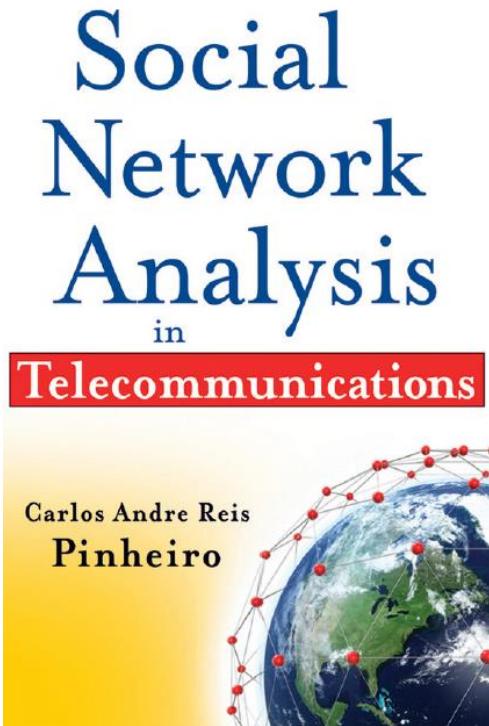
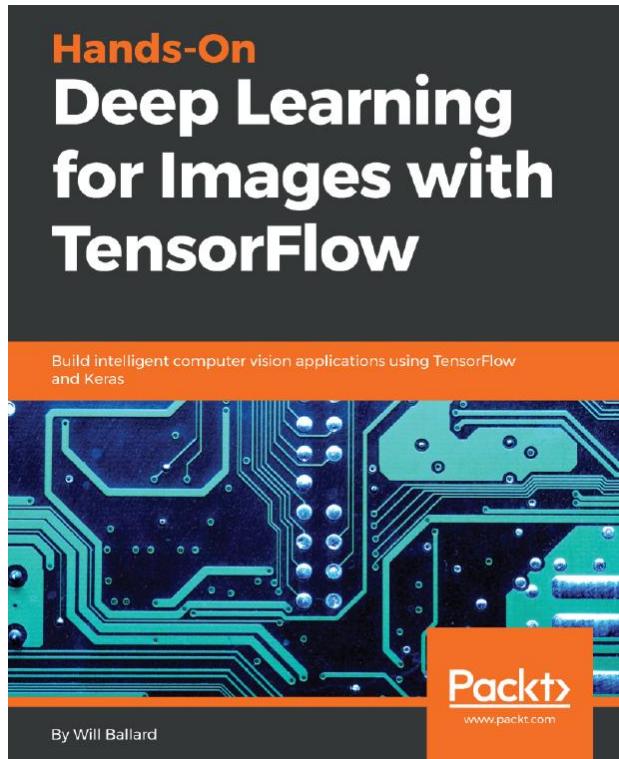
Colección para Principiantes



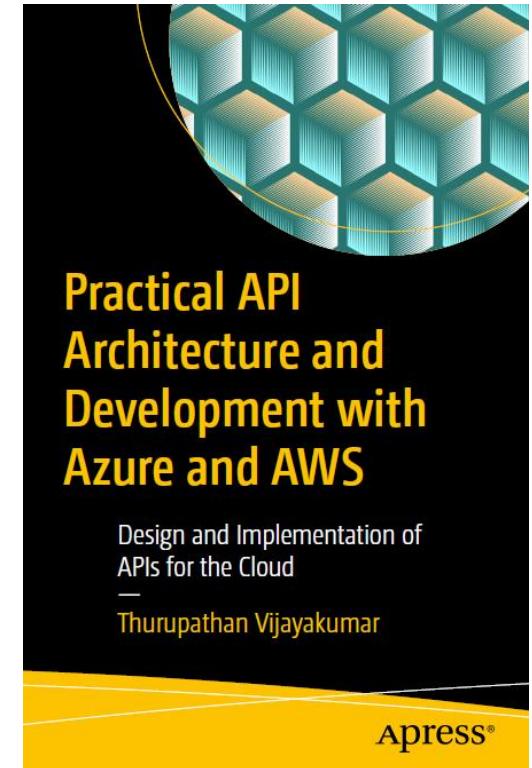
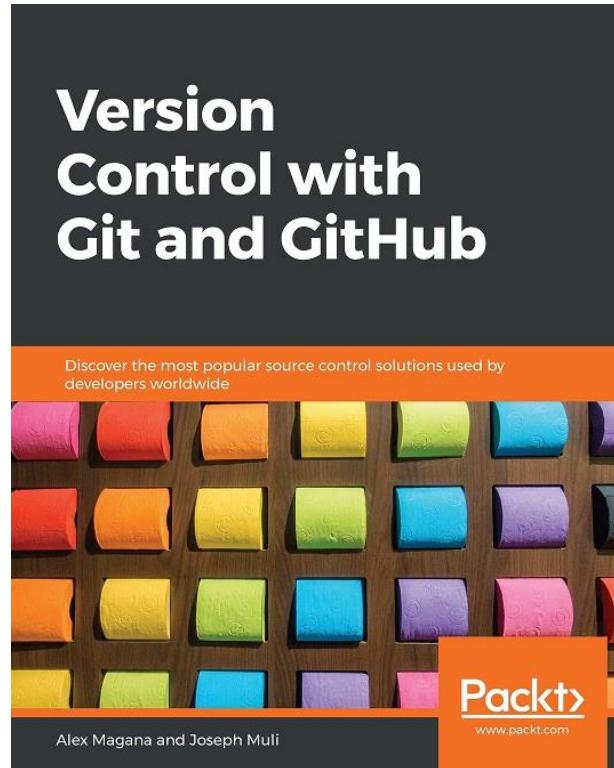
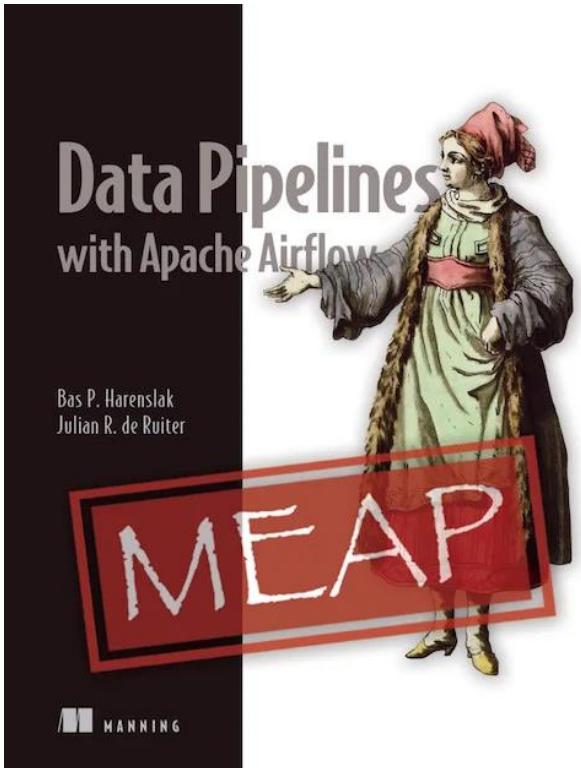
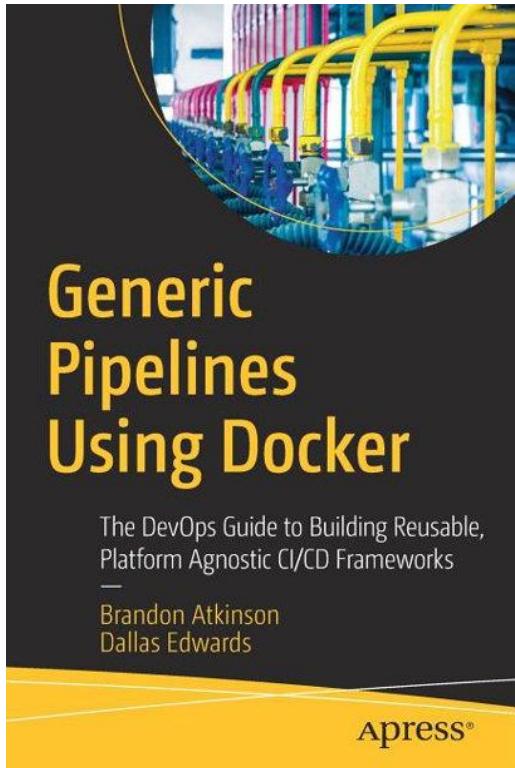
Colección Learn by Doing



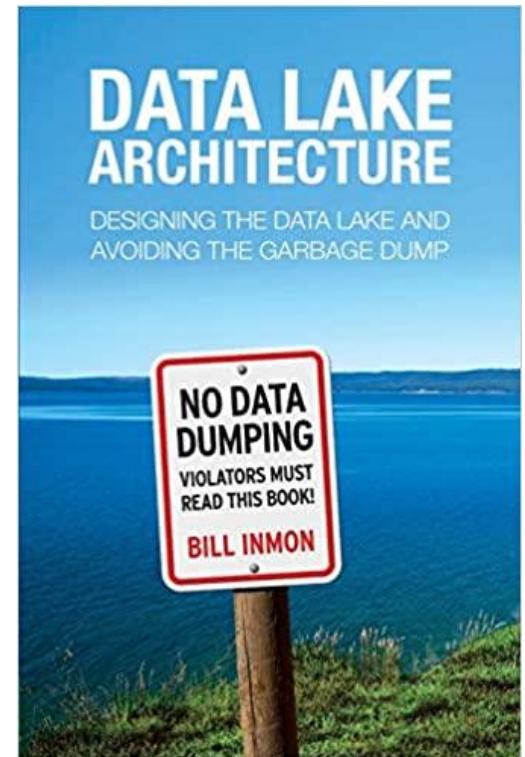
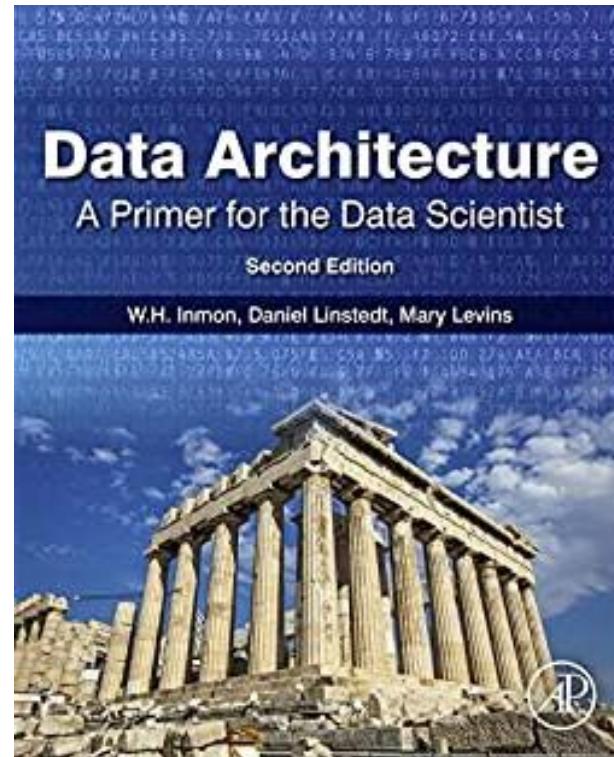
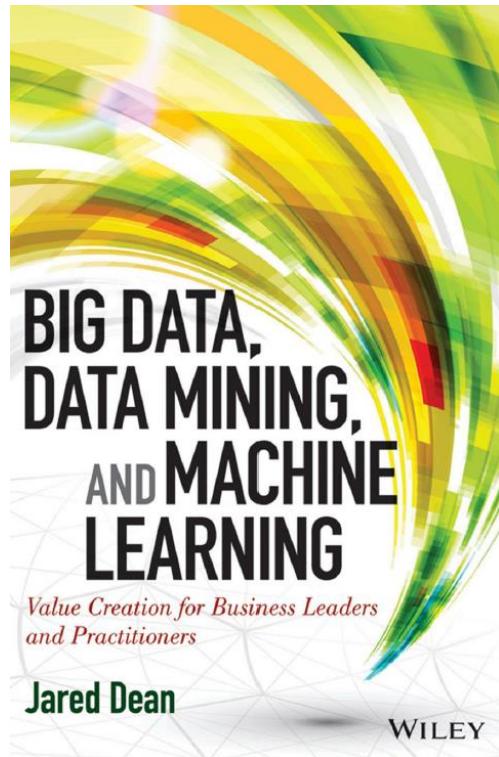
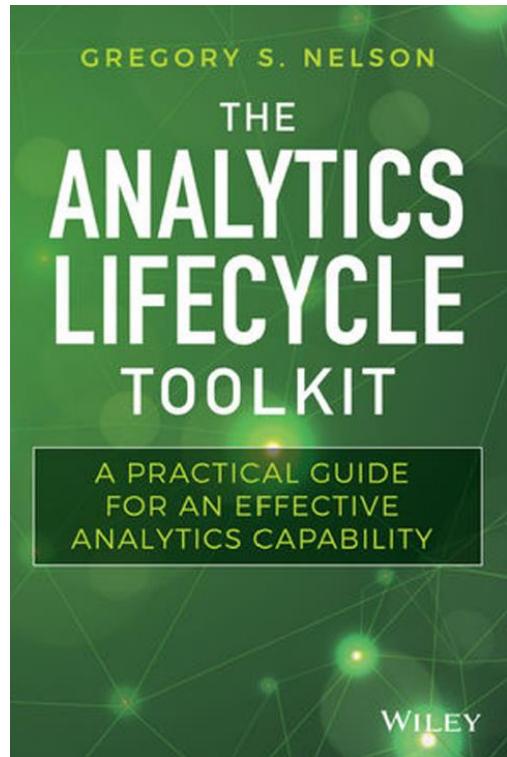
Colección Learn by Doing



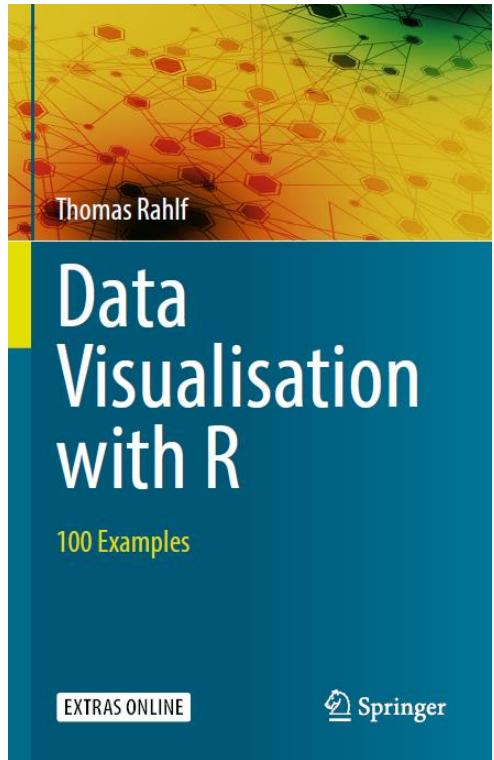
Colección para Data Engineers



Colección para Arquitectos

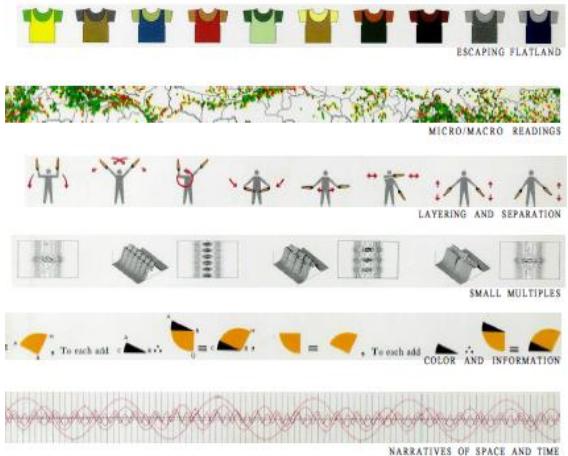


Colección para Data Artists



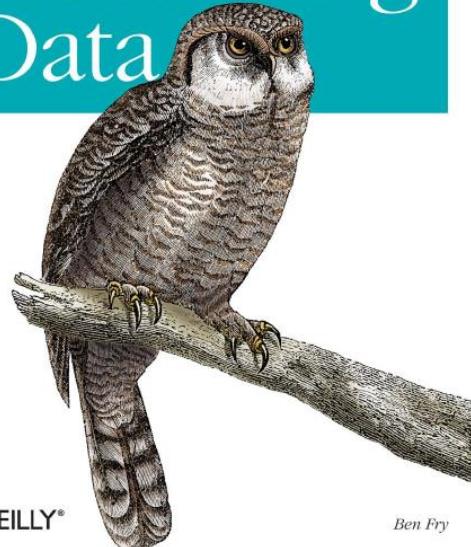
Edward R. Tufte

Envisioning Information

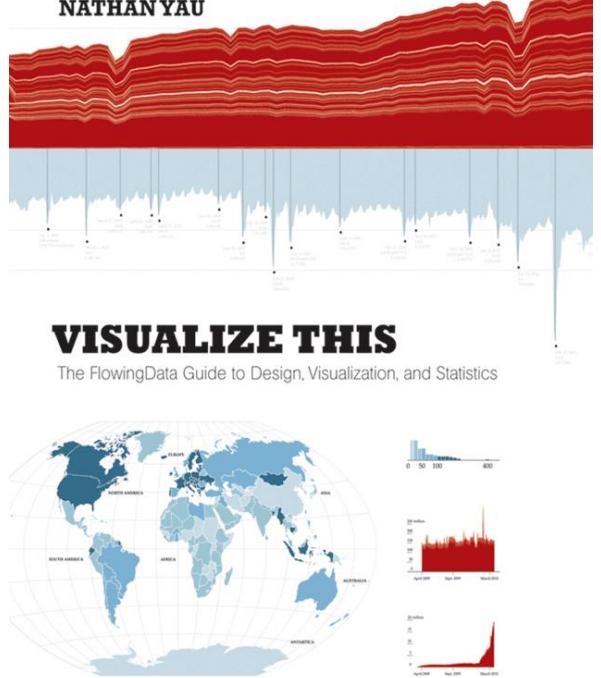


Exploring and Explaining Data with the Processing Environment

Visualizing Data

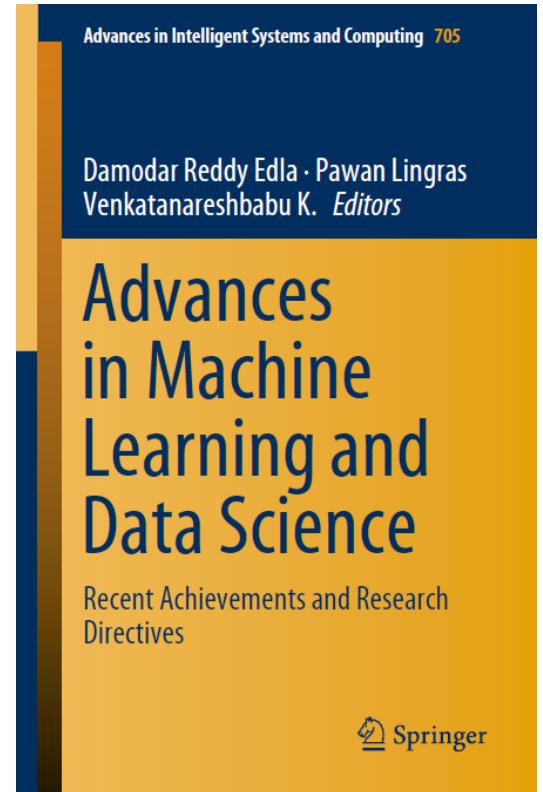
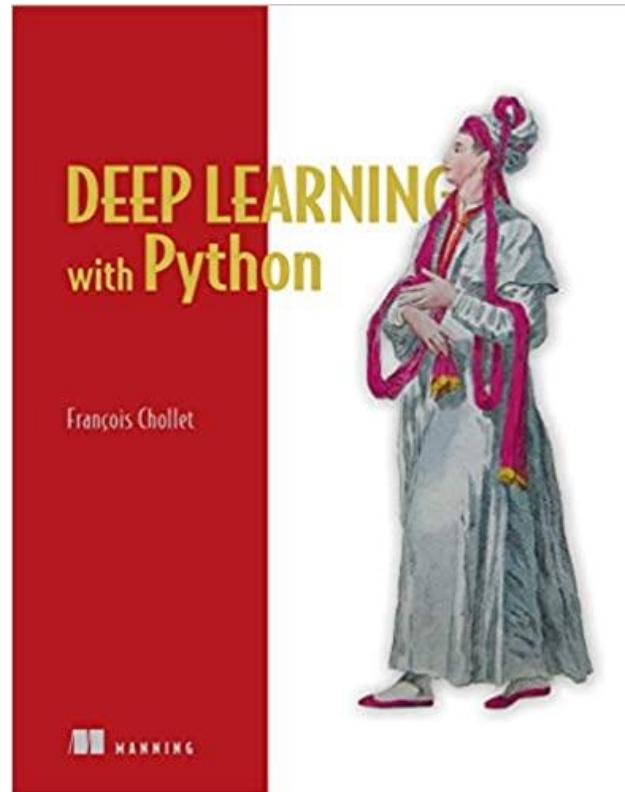
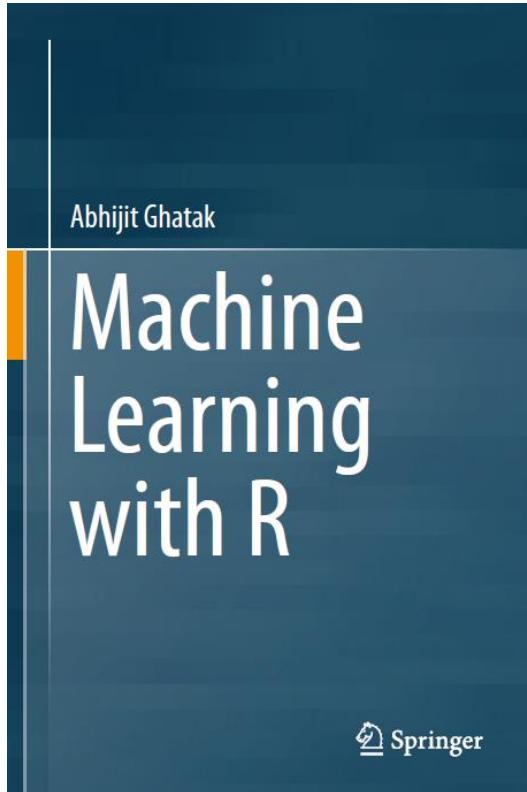
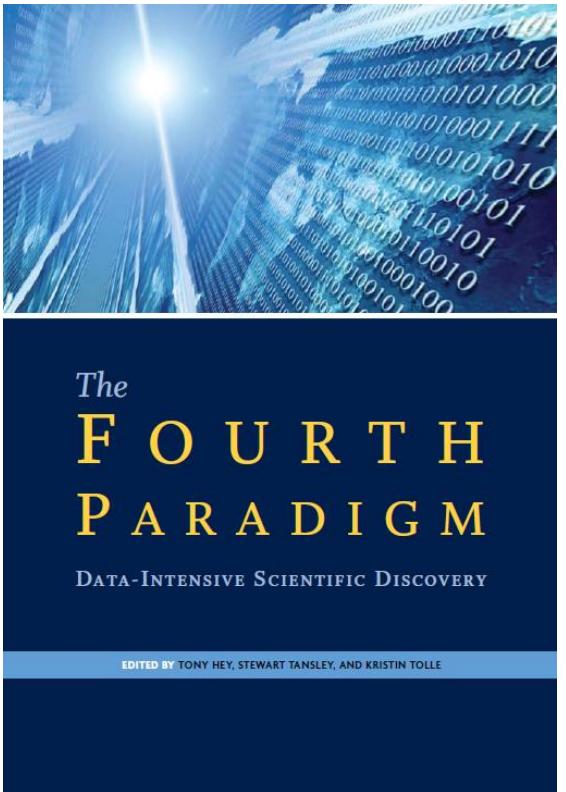


Ben Fry



teradata.

Colección para Researchers



Guías Metodológicas

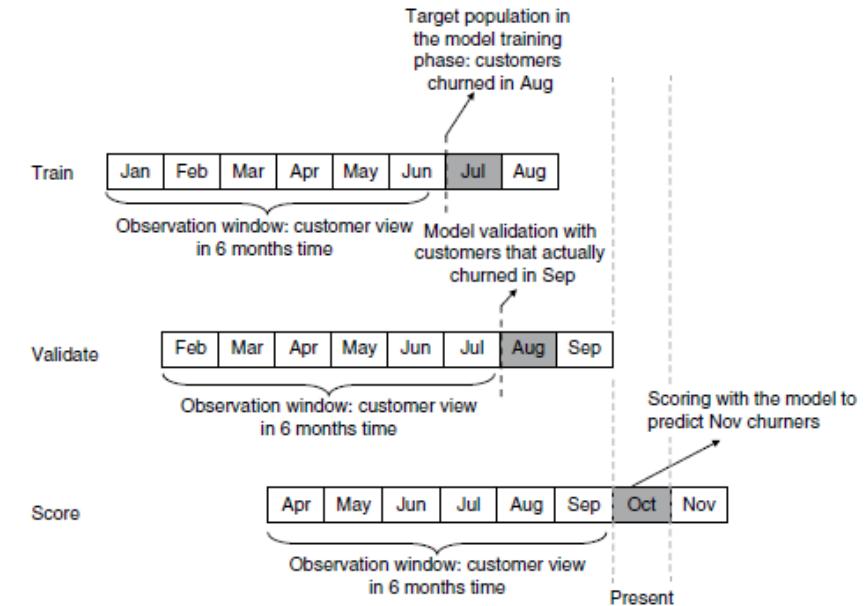
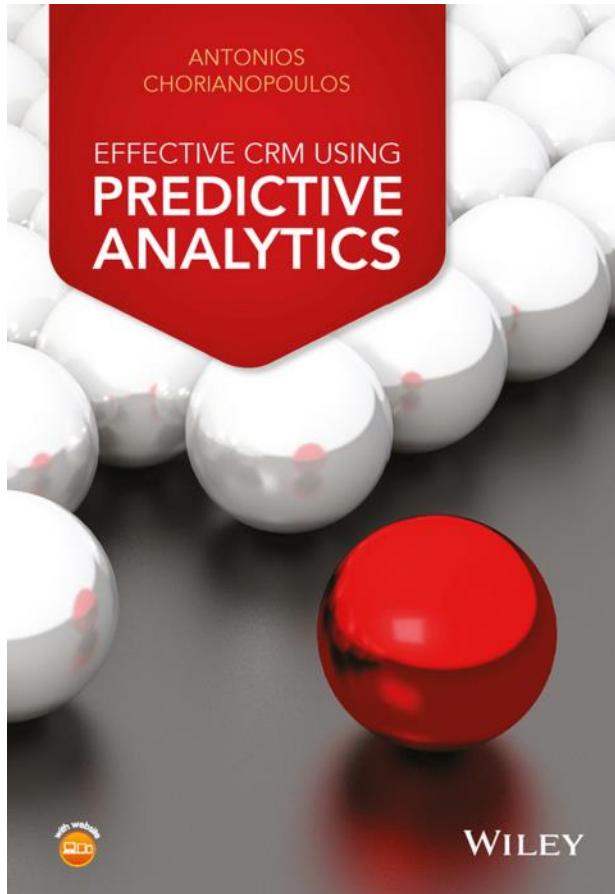
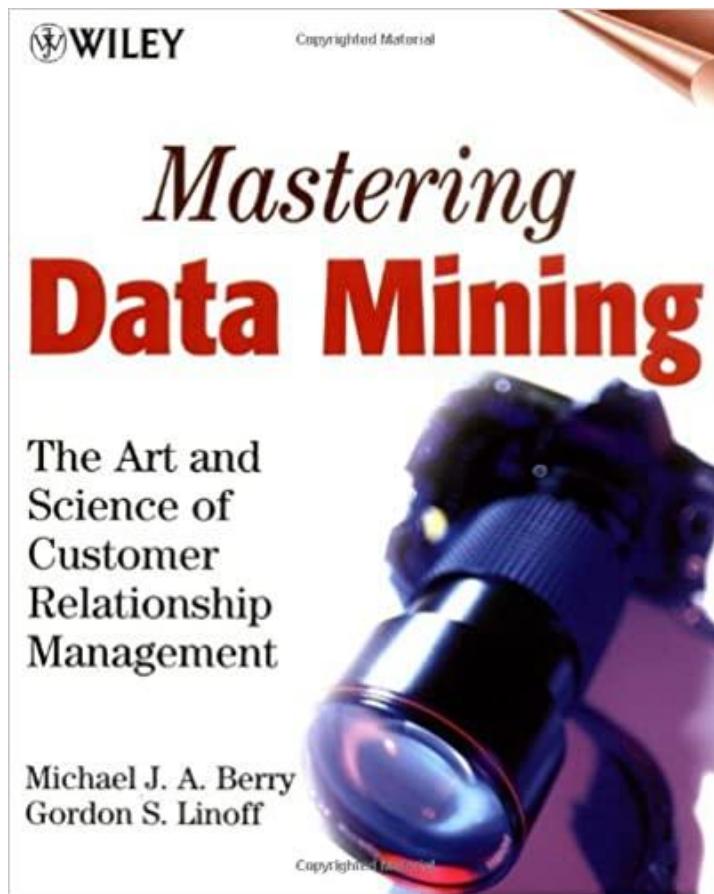
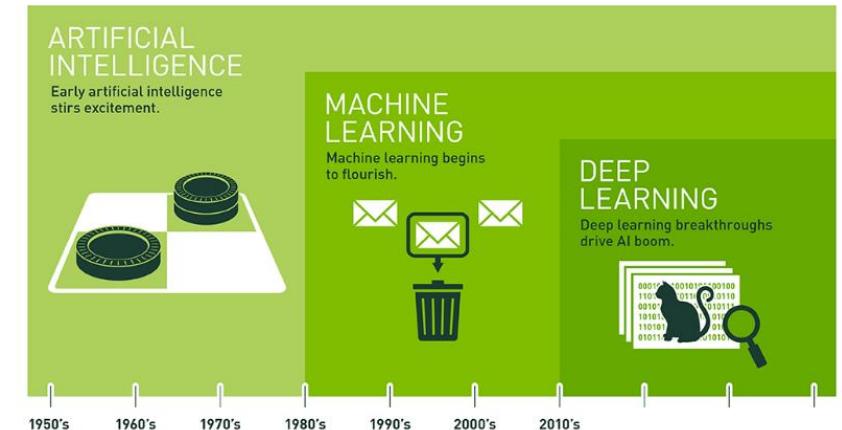
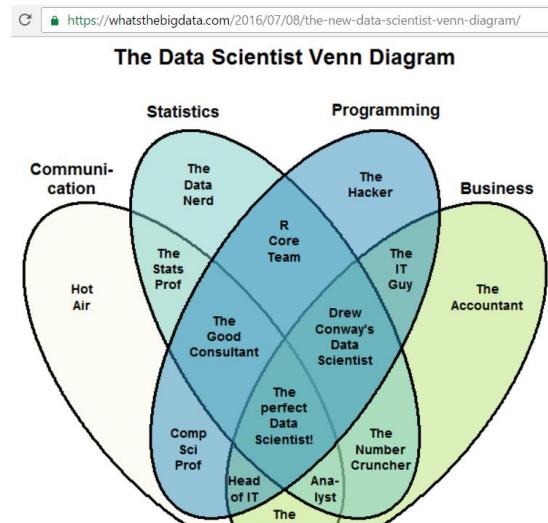
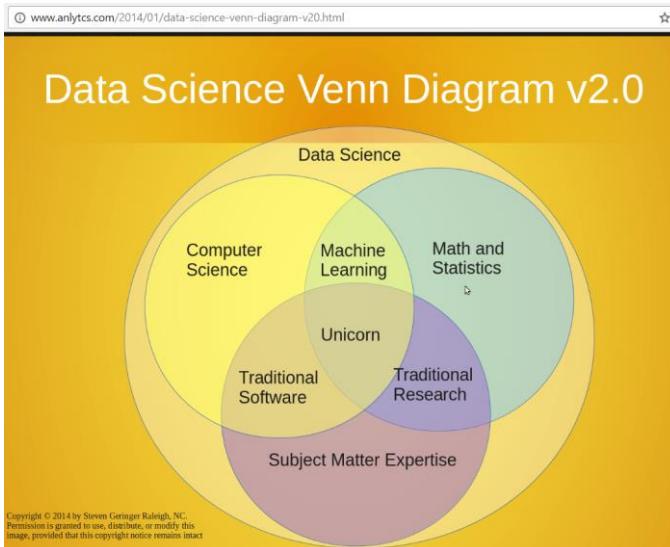
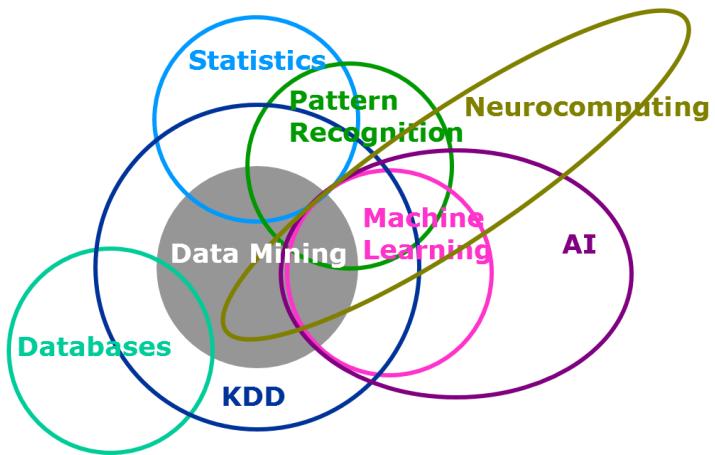


Figure 2.2 The data setup and time frames in a churn model. Source: Tsipitsis and Chorianopoulos (2009). Reproduced with permission from Wiley



Recomendaciones Específicas

Dejemos las discusiones improductivas



Lo más importante es entender el problema

- Primero, una historia (Septiembre 2015)



La primera competencia analítica
que se realizó en Perú

- Objetivo: Mejorar el algoritmo de matching (recomendación), para que los usuarios encuentren pareja más rápido.

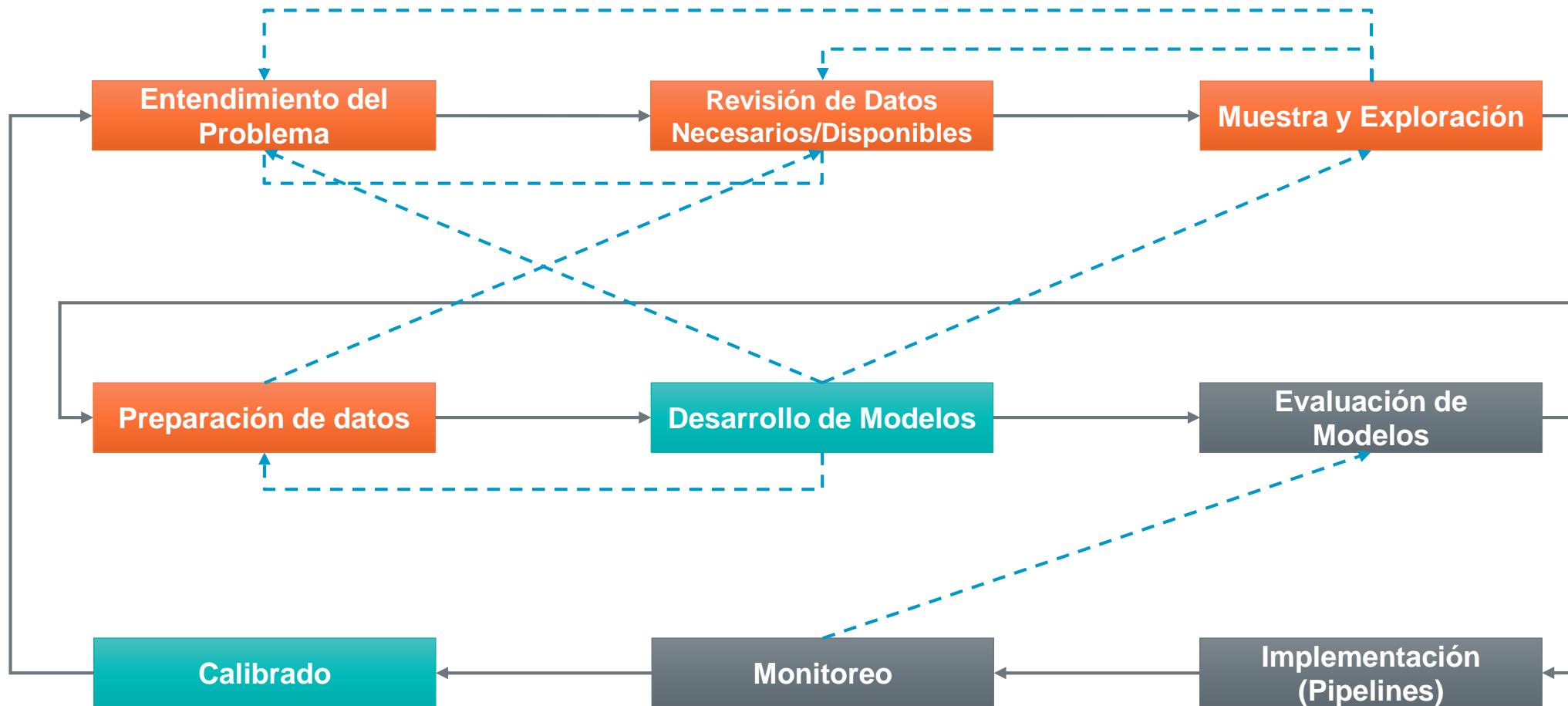
Lo más importante es entender muy bien el problema

- Y en junio de 2019:

The screenshot shows a news article from the website revistaganamas.com.pe. The header features the Gan@Más logo and navigation links for Emprendimientos, Innovación, Finanzas, Empresas, Economía, Negocios, Actualidad, Sociedad, and Blogs. The main headline reads "Mi Media Manzana cierra operaciones luego de 6 años en el mercado". Below the headline is a photo of a group of people at a bar. A sub-headline states "Pedro Neira, CEO de Mi Media Manzana, contó por qué la empresa cerró operaciones luego de permanecer en el mercado 6 años." At the bottom of the article, there is a note: "En un artículo publicado en su cuenta de LinkedIn Pedro explica que la razón principal del cierre es porque no lograron la rentabilidad para sus inversionistas." To the right of the article, there is a sidebar with a newsletter sign-up button and three other news snippets.

En un artículo publicado en su cuenta de LinkedIn Pedro explica que la razón principal del cierre es porque no lograron la rentabilidad para sus inversionistas.

Por eso va primero en el ciclo de vida



Pero, pongamos las cosas en perspectiva

Hidden Technical Debt in Machine Learning Systems

D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips
{dsculley, gholt, dg, edavydov, toddphillips}@google.com
Google, Inc.



Source: <https://papers.nips.cc/paper/5656-hidden-technical-debt-in-machine-learning-systems.pdf>

Una dosis de sentido común nunca está demás

☰ MENU

Harvard Business Review

DATA

If Your Data Is Bad, Your Machine Learning Tools Are Useless

by Thomas C. Redman

APRIL 02, 2018

SUMMARY | SAVE | SHARE | 10 COMMENT | HH TEXT SIZE | PRINT | \$8.95 BUY COPIES



1/3 FREE ARTICLES LEFT > REGISTER FOR MORE | SUBSCRIBE + SAVE! 🎉

☰ MENU

Harvard Business Review

MARKETING

Use Big Data to Create Value for Customers, Not Just Target Them

by Niraj Datar

AUGUST 16, 2016

SAVE | SHARE | 3 COMMENT | HH TEXT SIZE | PRINT | \$8.95 BUY COPIES



<https://hbr.org/2018/04/if-your-data-is-bad-your-machine-learning-tools-are-useless>

<https://hbr.org/2016/08/use-big-data-to-create-value-for-customers-not-just-target-them>



<https://www.enriquedans.com/2006/08/panales-y-cerveza.html>

20 agosto 2006

En beer, correlation, customer intelligence, diapers, market basket analysis, statistics

¿Pañales y cerveza?

Hasta aquí, todo muy bien. Es, como dice [Seth Godin en su fantástico libro](#), una buena historia, una de esas que aparentemente tiene patas, que se sostiene, que gusta contar. ¿El problema? Muy sencillo: [que es mentira](#). Lisa y llanamente falsa. Es una recolección de datos inconexos que comienza con el descubrimiento de una inofensiva correlación entre consumo de pañales y cerveza no adscrita a sexo ni a día de la semana alguno, que continúa con unos cambios en el *store layout* que las grandes superficies nunca tomaron, y que prosigue con los supuestos buenos resultados que las grandes superficies presuntamente obtuvieron. En realidad, nada. Simplemente, una correlación inofensiva no sustentada que podría servir para ilustrar la diferencia entre la búsqueda orientada de hipótesis y la mera exploración de los datos, y que los vendedores de software de BI suelen usar con profusión para poner los dientes largos a los clientes con eso de "anda que si encuentras algo así, inesperado y rentable..." Pues a lo mejor lo encuentras, pero esperemos que si lo haces sea, al menos, más cierto que la historia de los pañales y la cerveza.

Buscador por temas

Enrique Dans

Curriculum
Contacto
Mi libro: "Todo va a cambiar"
En Google+
En Facebook
Artículos antiguos
Enlaces en Pinterest
Mis fotos en Flickr

Hasta aquí, todo muy bien. Es, como dice [Seth Godin en su fantástico libro](#), una buena historia, una de esas que aparentemente tiene patas, que se sostiene, que gusta contar. ¿El problema? Muy sencillo: [que es mentira](#). Lisa y llanamente falsa. Es una recolección de datos inconexos que comienza con el descubrimiento de una inofensiva correlación entre consumo de pañales y cerveza no adscrita a sexo ni a día de la semana alguno, que continúa con unos cambios en el *store layout* que las grandes superficies nunca tomaron, y que prosigue con los supuestos buenos resultados que las grandes superficies presuntamente obtuvieron. En realidad, nada. Simplemente, una correlación inofensiva no sustentada que podría servir para ilustrar la diferencia entre la búsqueda orientada de hipótesis y la mera exploración de los datos, y que los vendedores de software de BI suelen usar con profusión para poner los dientes largos a los clientes con eso de "anda que si encuentras algo así, inesperado y rentable..." Pues a lo mejor lo encuentras, pero esperemos que si lo haces sea, al menos, más cierto que la historia de los pañales y la cerveza.

Para más explicaciones, este [artículo en The Register](#) y esta [anotación de Daniel J. Power](#), uno de los implicados en el comienzo de esta leyenda urbana.

En la vida real...

- ¿Si compra PAÑALES, es probable que compre CERVEZAS?



	col1_item1	col1_item2	cntb	cnt1	cnt2	score	support	confidence	lift
1	PAÑAL BEBE	CERVEZA IMPORTADA	2514	570328	140709	7.875605...	7.5620282...	0.0044079...	1.0414...
2	PAÑAL BEBE	CERVEZA NACIONAL	8683	570328	611507	2.161789...	2.6118174...	0.0152245...	0.8276...

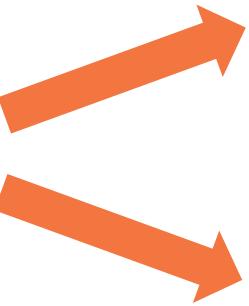
- ¿Si compra CERVEZAS, es probable que compre PAÑALES?

	col1_item1	col1_item2	cntb	cnt1	cnt2	score	support	confidence	lift
1	CERVEZA IMPORTADA	PAÑAL BEBE	2514	140709	570328	7.875605...	7.5620282...	0.01786666...	1.0414...
2	CERVEZA NACIONAL	PAÑAL BEBE	8683	611507	570328	2.161789...	2.6118174...	0.01419934...	0.8276...

Afinidades Significativas



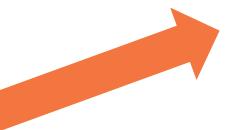
Lift = 2.2



Lift = 1.4



Lift = 4.6



Lift = 3.2



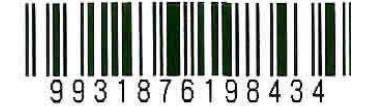
col1_item1	col1_item2	cntb	cnt1	cnt2	score	support	confidence	lift
TINTO	VARIOS	4231	7705685624	0.00271321...	0.0011195...	0.0549081...	2.4235...	
TINTO	HIELO	980	7705619845	6.28050531...	2.5930508...	0.0127180...	2.4220...	
TINTO	ENCURTIDOS	4526	7705691718	0.00289846...	0.0011975...	0.0587365...	2.4202...	
TINTO	HONGOS	3264	7705668506	0.00201820...	8.6364468...	0.0423588...	2.3368...	
TINTO	SIN HUESO	713	7705615298	4.31258772...	1.8865767...	0.0092530...	2.2859...	
TINTO	ENJUAGUE BUCAL	650	7705614335	3.82492175...	1.7198806...	0.0084354...	2.2239...	
TINTO	DULCES	2768	7705661277	0.00162266...	7.3240456...	0.0359219...	2.2155...	
TINTO	ORGANIZADORES ...	36	77056799	2.10499835...	9.5254928...	4.6719269...	2.2098...	
TINTO	PAPELERIA	2	7705645	1.15356220...	5.2919404...	2.5955149...	2.1798...	
TINTO	AL HUEVO	429	770569688	2.46532394...	1.1351212...	0.0055673...	2.1718...	

col1_item1	col1_item2	cntb	cnt1	cnt2	score	support	confidence	lift
TINTO	PISCO	2606	7705620340	0.00433302...	6.8953984...	0.038195...	6.2839...	
TINTO	CERVEZA IMPORTA...	2552	7705630430	0.00277749...	6.7525160...	0.0331187...	4.1132...	
TINTO	BOSQUE	2456	7705668409	0.00114429...	6.4985029...	0.0318729...	1.7608...	
TINTO	ESPECIAL	2410	77056170646	4.41704182...	6.3767882...	0.0312759...	0.6926...	
TINTO	JABON	2409	7705681929	9.19241208...	6.3741423...	0.0312629...	1.4421...	
TINTO	ESPECIAS	2292	7705663350	0.00107615...	6.0645637...	0.0297446...	1.7745...	
TINTO	GALLETAS INTEGRA...	2285	7705688294	7.67422902...	6.0460419...	0.0296537...	1.2692...	
TINTO	COMIDAS	2260	77056203881	3.25112496...	5.9798927...	0.0293293...	0.5436...	
TINTO	PREPARADOS	2245	7705659614	0.00109718...	5.9402031...	0.0291346...	1.8470...	

col1_item1	col1_item2	cntb	cnt1	cnt2	score	support	confidence	lift
PAPEL HIGIENICO	TALCO	425	963383566	5.257737...	1.12453735...	0.004411550...	4.675467...	
PAPEL HIGIENICO	ACCESORIOS DE BAÑO	328	963382786	4.008380...	8.67878238...	0.003404679...	4.618597...	
PAPEL HIGIENICO	SAL	3540	9633830089	0.004323...	9.36673465...	0.036745624...	4.615438...	
PAPEL HIGIENICO	VINAGRE	2250	9633819127	0.002747...	5.95343304...	0.023355269...	4.614802...	
PAPEL HIGIENICO	ATUN	9362	9633880181	0.011346...	0.00247715...	0.097178683...	4.580517...	
PAPEL HIGIENICO	CEPILLO DENTAL	2070	9633817833	0.002494...	5.47715839...	0.021486848...	4.553688...	
PAPEL HIGIENICO	CUIDADO CORPORAL	676	963385833	8.132120...	1.78867588...	0.007016961...	4.546447...	
PAPEL HIGIENICO	ENVASADA	17196	96338151232	0.020296...	0.00455001...	0.178496543...	4.460680...	
PAPEL HIGIENICO	DESODORANTE HOMBRE	3202	9633828347	0.003754...	8.47239670...	0.033237144...	4.431304...	

col1_item1	col1_item2	cntb	cnt1	cnt2	score	support	confidence	lift
PAPEL HIGIENICO	PAN	19236	96338237901	0.016144...	0.00508978...	0.199671988...	3.172020...	
PAPEL HIGIENICO	TROZADO	18799	96338233405	0.015716...	0.00497415...	0.195135875...	3.159672...	
PAPEL HIGIENICO	JAMON COCIDOS	17608	96338313791	0.010256...	0.00465902...	0.182773152...	2.201339...	
PAPEL HIGIENICO	ENVASADA	17196	96338151232	0.020296...	0.00455001...	0.178496543...	4.460680...	
PAPEL HIGIENICO	SIN GAS	16901	96338324172	0.009146...	0.00447195...	0.175434408...	2.045287...	
PAPEL HIGIENICO	SEMI-DUROS	16321	96338263591	0.010489...	0.00431848...	0.169413938...	2.429034...	
PAPEL HIGIENICO	JABON	16109	9633881929	0.032877...	0.00426239...	0.167213352...	7.713444...	
PAPEL HIGIENICO	GRANEL	15999	96338205384	0.012936...	0.00423328...	0.166071539...	3.055931...	
PAPEL HIGIENICO	UHT	15683	96338171845	0.014856...	0.00414967...	0.162791421...	3.580219...	

Accionamiento de los Insights



Nº de tarjeta [REDACTED]

10€ de DTO
en la sección de MASCOTAS
de tu hiper habitual

VÁLIDO EN HIPERMERCADOS CARREFOUR

Descuento válido hasta el 30-11-2011

UTILÍCELLO EN SU PRÓXIMA COMPRA CON SU TARJETA CLUB O PASS. NO SE PODRÁN CANJEAR 2 CUPONES IGUALES EN UN MISMO TICKET NI USAR EL MISMO CUPÓN MÁS DE UNA VEZ. CONSULTAR BASES DE EL CLUB

03/09/11 13:12:48 . 0170 018 0049 8793

teradata.

Key Takeaways

1

Las personas y procesos
son lo más importante en
una cultura data driven.

**Seamos parte de la
solución.**

2

Aprender del fracaso
(propio o ajeno) nos
ayudará a no tropezar
siempre con la misma
piedra.

3

Para aprender, más
importante que
encontrar un buen
maestro es **aprender
a buscar en el lugar
correcto.**

Gracias!

teradata.

©2020 Teradata