



UNIVERSIDAD  
PANAMERICANA

# Tecnologías de la Información

**Dr. Leon Felipe Palafox Novack**  
**[lpalafox@up.edu.mx](mailto:lpalafox@up.edu.mx)**

1

## Anuncios parroquiales

Por que todos queremos saber como  
se va a calificar

“ *A mi no me importan las buenas calificaciones, mis creaciones seran el testamento de mi desempeño.*

*Helmut Kohl*



## Objetivo

En esta clase estudiaremos el ciclo de vida de un proyecto de ciencia de datos, desde el análisis del negocio, recopilación de datos, creación de base de datos, desarrollo del modelo y despliegue del modelo.

## Outcomes



- Los estudiantes serán capaces de analizar un negocio, identificar los datos requeridos para ofrecer una solución óptima, seleccionar el modelo adecuado para ofrecer dicha solución y desplegar el modelo de acuerdo a las necesidades del problema.

# Requisitos



- Fundamentos de programación
- Bases de Datos
- Fundamentos de desarrollo web
- Principios de Negocios

# Calificación



La evaluación consistirá en:

- El proyecto final será el 40% de la evaluación final.
- El restante 60% será distribuido de la siguiente forma:
  - ▷ Dos exámenes parciales y un final
  - ▷ Cuatro tareas.
  - ▷ Participación en clase

# Proyecto Final



## ■ Una solución completa de ciencia de datos:

- ▷ Entendimiento del negocio
- ▷ Recopilación de los datos
  - ▷ Entendimiento de los datos
  - ▷ Creación de ETL
- ▷ Creación del modelo
  - ▷ Pruebas, prototipos
- ▷ Puesta en marcha de la solución



## CONCEPTOS BÁSICOS

- ▷ Introducción a la Ciencia de Datos
  - ▷ Aplicaciones y soluciones
  - ▷ Elementos de la ciencia de datos
    - ▷ Machine Learning, Big Data, Business Intelligence
  - ▷ Definiciones básicas.
- ▷ Ciclo de vida de un proyecto de ciencia de datos

# Temario



- Entendimiento del negocio
  - ▷ Creación de Valor
    - ▷ Cash Flow, Revenue
    - ▷ Análisis de Costos
  - ▷ Entendimiento del mercado
    - ▷ Oferta, competencia
  - ▷ Entendimiento del problema

## ■ RECOPIACIÓN DE DATOS Y CENTRALIZACIÓN

- ▷ Bases de datos centralizadas.
  - ▷ Data Warehouse, Data Lake
- ▷ Silos de datos
- ▷ Extract, Transform and Load
- ▷ Limpieza de los datos
- ▷ Discriminación de datos

## ■ Exploración y presentación

- ▷ Análisis exploratorio de los datos
  - ▷ Tipos de gráfica
  - ▷ Estadísticas básicas
- ▷ Presentación de datos y resultados
  - ▷ Storytelling
  - ▷ Presentación de gráficas

## ■ Modelos de Aprendizaje Máquina

- ▷ Introducción
  - ▷ Algoritmos Supervisados y no Supervisados
  - ▷ Clasificadores
  - ▷ Reducción de dimensionalidades
- ▷ Procesamiento de Lenguaje Natural
- ▷ Modelos Bayesianos

## ■ Implementación

- ▷ Pruebas de validación
- ▷ Modelos en tiempo real
- ▷ Pruebas de stress
  - ▷ Calendarios de re-entrenamiento
- ▷ Big Data (?)

## Libro de Texto

- Bishop, Christopher M. Pattern recognition and machine learning. Springer, 2006.
- Rogers, Simon, and Mark Girolami. A first course in machine learning. CRC Press, 2011.  
(<http://www.dcs.gla.ac.uk/~srogers/firstcourseml/>)
- James, Gareth, et al. An introduction to statistical learning. New York: Springer, 2013. (<http://www-bcf.usc.edu/~gareth/ISL/>)

# 2

## Introducción

Que es la Ciencia de Datos



# ¿Qué es la ciencia de datos?



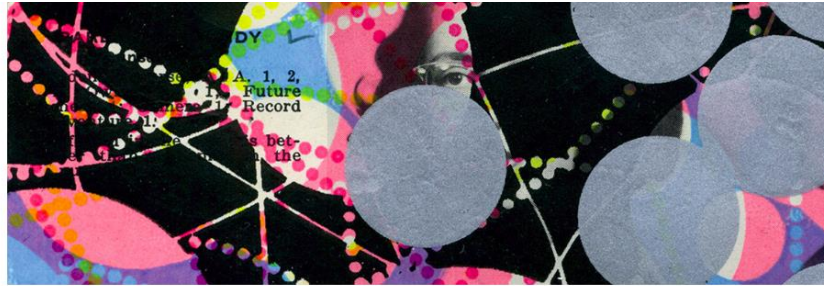
## ■ Que es un Data Scientist?

- ▷ No es un analista
- ▷ No es un programador
- ▷ No es un programador de base de datos.

## ■ Hace Ciencia?

- ▷ No nos importa

# Historia de la Ciencia de Datos



DATA

## Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

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



When Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing

# Historia de la Ciencia de Datos

## TECHNOLOGY

### These data scientists are disrupting Disney World's long wait times

*Touring Plans* goes to impressive lengths to help you avoid ride lines.

By Carlye Wise | October 31, 2017



### The Incredible Ways Shell Uses Artificial Intelligence To Help Transform The Oil And Gas Giant

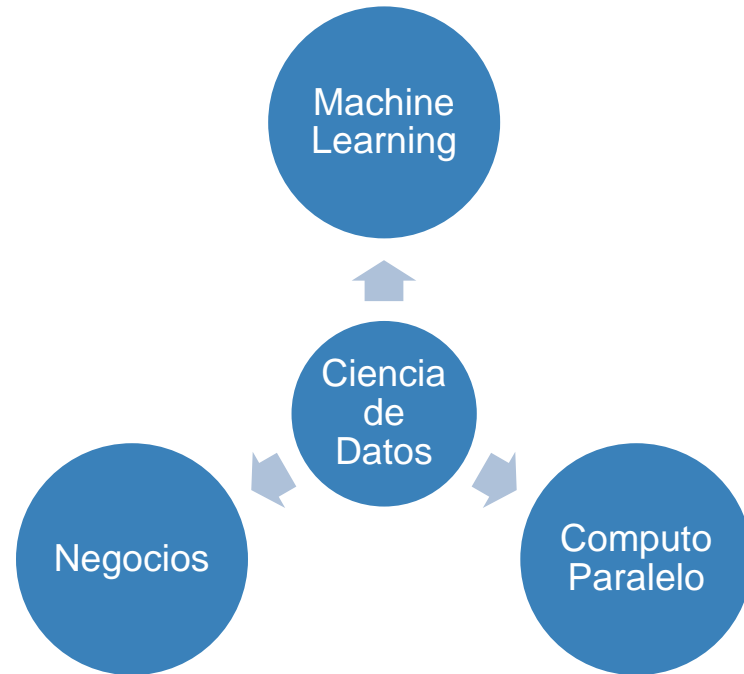


**Bernard Marr** Contributor  
Enterprise & Cloud

Royal Dutch Shell is heavily investing in research and development of [artificial intelligence \(AI\)](#), which it hopes will provide solutions to some of its most pressing challenges.

From meeting the demands of a transitioning energy market, urgently in need of cleaner and more efficient power, to improving safety on the forecourts of its service stations, AI is at the top of the agenda. I have been working with Shell over the past months to help create a data strategy, which gave me a thorough insight into Shell's AI priorities and initiatives.

# Historia de la Ciencia de Datos



# Historia de la Inteligencia Artificial

1960

- Nacimiento
- Perceptron

1980

- Redes Neuronales
- Mucho Dinero
- Muchas Promesas

1990

- El invierno de la AI
- Se perdió dinero
- Rebranding

2000  
s

- Machine Learning
- Aprendizaje Estadístico

2010

Deep Learning

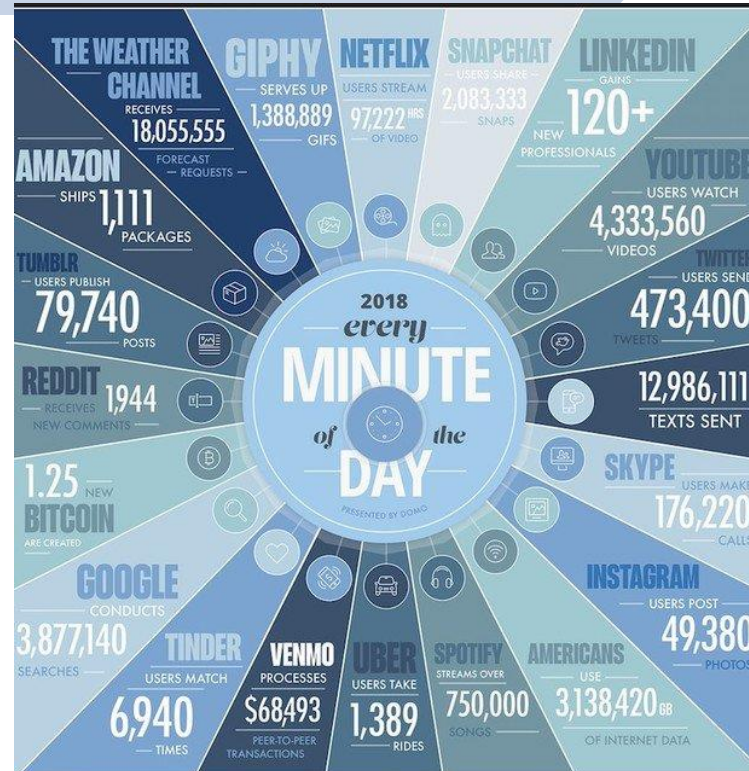
- Faltaban dos cosas:
  - ▷ Poder de procesamiento
  - ▷ Datos!

# Big Data





UNIVERSIDAD  
PANAMERICANA





## Procesar datos masivos para obtener:

- Accionables
- Información
- Ventaja



## Perfilamiento de Clientes

- Se necesitan datos de los clientes
- Se necesita saber las tendencias de compra



## Datos de los clientes:

- ▶ Tarjetas de puntos
- ▶ Tarjetas de crédito
- ▶ Tarjetas de lealtad
- ▶ Compra en línea.

Que productos compró

Que tiendas visitó

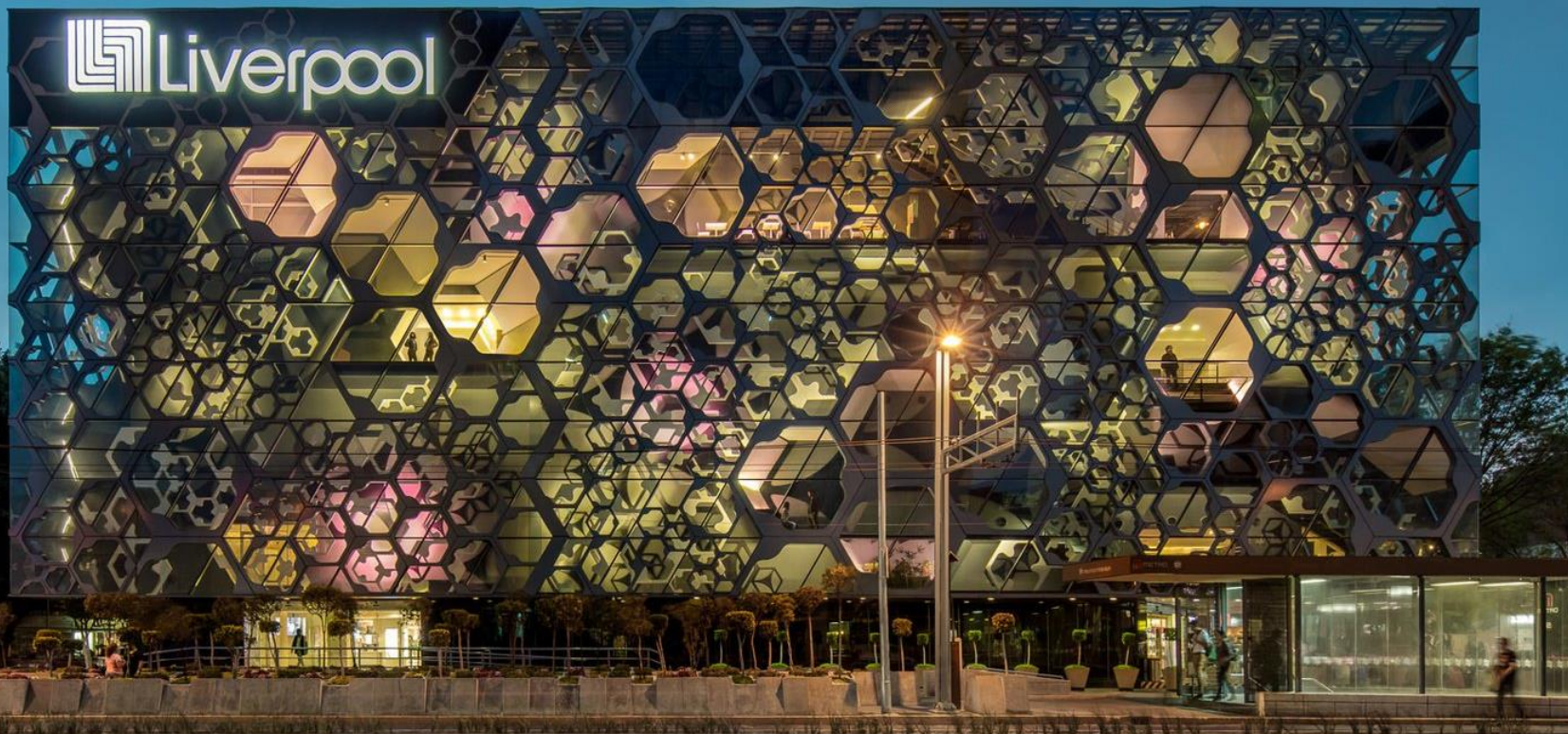
Que productos vio



amazon







# Perfilamiento de Clientes



- Algoritmos de recomendación de productos
- Algoritmos de recomendación de promociones.
- Algoritmos de recomendación de experiencias.
  - ▶ Marketing 4.0



# Nuevos puntos de ventas

- Donde ponemos los nuevos puntos de ventas.
- Queremos maximizar las ganancias.

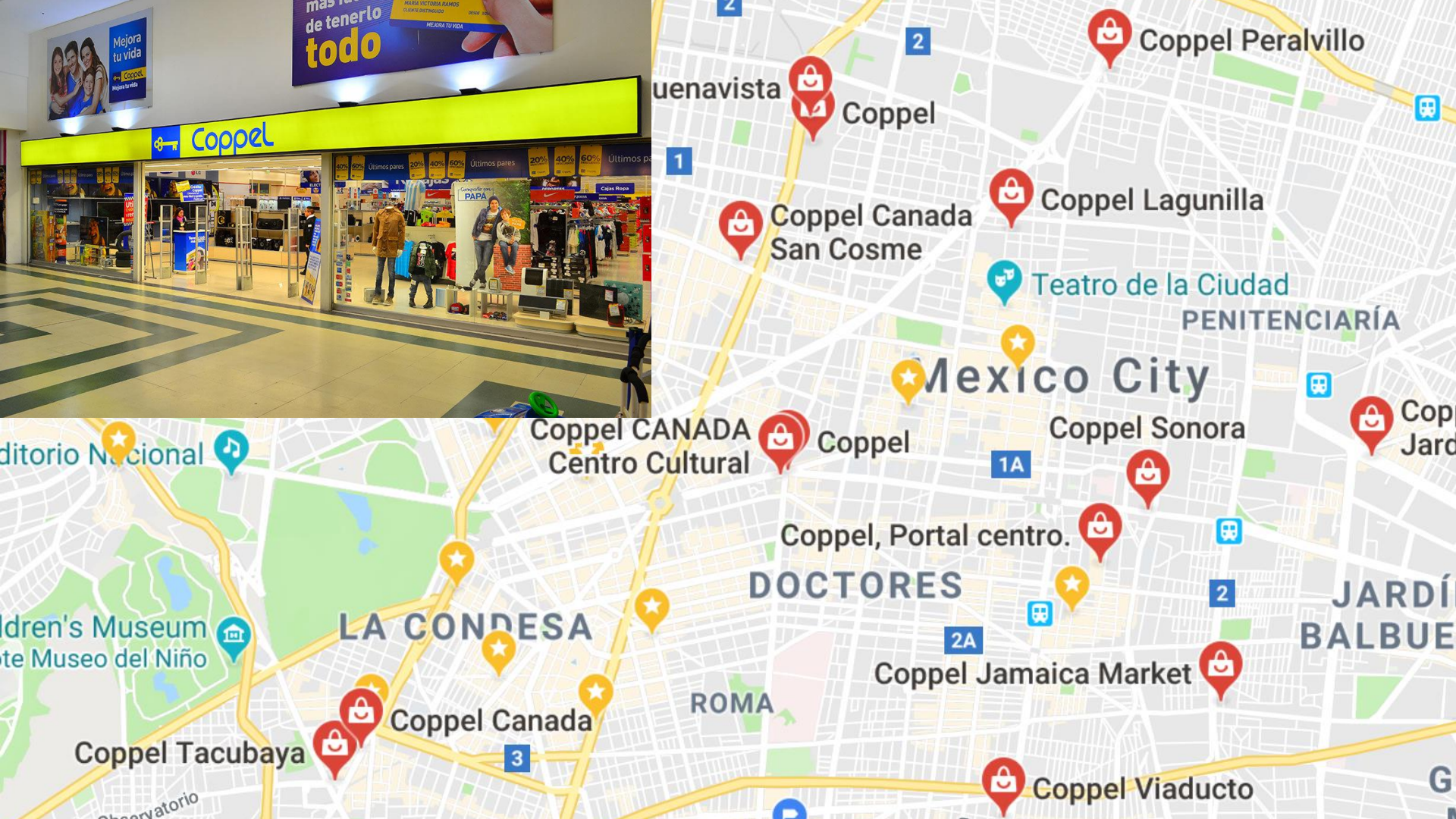




Se tiene que tener un registro claro de ventas.

- ▶ Clientes, productos, servicios.
- ▶ Volúmenes de ventas.
- ▶ Área de las tiendas.





# Nuevos puntos de venta



- Algoritmos de detección de similitudes
- Algoritmos de recomendación de productos
- Algoritmos de coordinación geográfica.

