
Minería de Datos I

Presentación del Curso

Objetivos

Teóricos

- Identificar y aplicar las etapas esenciales del proceso de Minería de Datos en conjuntos de datos reales.
- Identificar patrones, relaciones y valores atípicos en conjuntos de datos.
- Evaluar y comparar el rendimiento de diferentes modelos utilizando métricas relevantes.
- Aplicar algoritmos de agrupamiento para segmentar datos, además de realizar reducción de dimensionalidad y poder interpretar los resultados obtenidos.

Prácticos

- Analizar y clasificar textos, realizar preprocesamiento de texto y extracción de características, y explorar las aplicaciones de Minería de Datos en el procesamiento de texto.
- Aplicar técnicas de preprocesamiento y extracción de características en imágenes, y explorar aplicaciones de Minería de Datos en imágenes y visión por computadora.
- Identificar aplicaciones prácticas de la Minería de Datos en diversos campos, evaluar las implicaciones éticas y de privacidad, y utilizar los resultados de manera responsable y consciente.

Planificación

En el siguiente link se encuentra la **planificación tentativa del curso** y **referencias bibliográficas**:



Requisitos previos



Programación: Los estudiantes deben estar cómodos con al menos un lenguaje de programación, preferiblemente Python.

Estadísticas: Una comprensión básica de conceptos estadísticos como media, mediana, desviación estándar y distribuciones probabilísticas.



Matemáticas: Conocimientos fundamentales de matemáticas, incluyendo álgebra lineal y cálculo diferencial.

Bases de Datos: Comprender conceptos de bases de datos, como tablas, registros, claves primarias y consultas SQL.



Evaluación

Curso no promocionable.

- Participación en clase.
- Dos exámenes parciales (04-10-23 y 29-11-23):
 - Se aprueban con 4 o más.
 - Son de aprobación obligatoria.
 - Tienen recuperatorio (06-12-23)
- Dos trabajos prácticos (grupales, por definir):
 - Se aprueban con 4 o más.
 - Son de aprobación obligatoria.
 - No se recuperan.
- Examen final
 - Se aprueba con 4 o más
 - Tiene dos instancias (13-12-23 y 20-12-23)

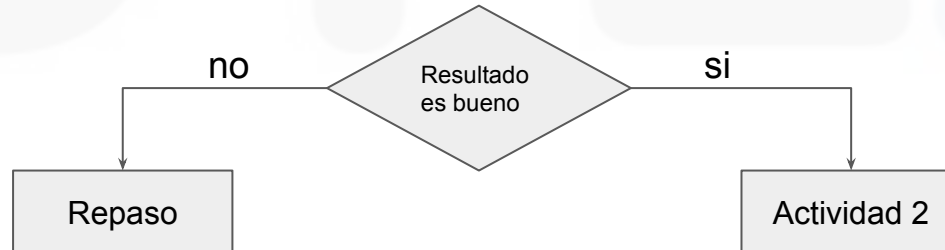
Midamos un poco...

Actividad:

Vamos a medir algo de conocimientos actuales:



<https://www.menti.com/>



Caso introductorio 1:

¿Alguien conoce el caso de las
cervezas y los pañales?



Caso introductorio 1: Walmart (caso emblemático)

Se refiere a un análisis realizado por Walmart en el que se **descubrió** una **relación sorprendente** entre la compra de *cerveza* y *pañales* en las tiendas.

“estadísticamente significativa”



Caso introductorio 1: Walmart (caso emblemático)

Patrón:

Padres jóvenes que después de un día de trabajo, compraban cervezas y pañales en la misma visita a la tienda



Caso introductorio 2: El juego de hoy en la NBA

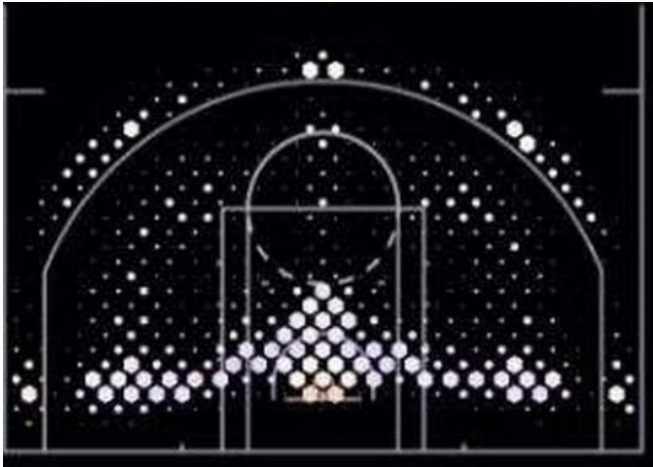
¿Alguien sabe cómo era el juego en la NBA en el año 2000?



¿Por qué el juego hoy es tan diferente?

Caso introductorio 2: El juego de hoy en la NBA

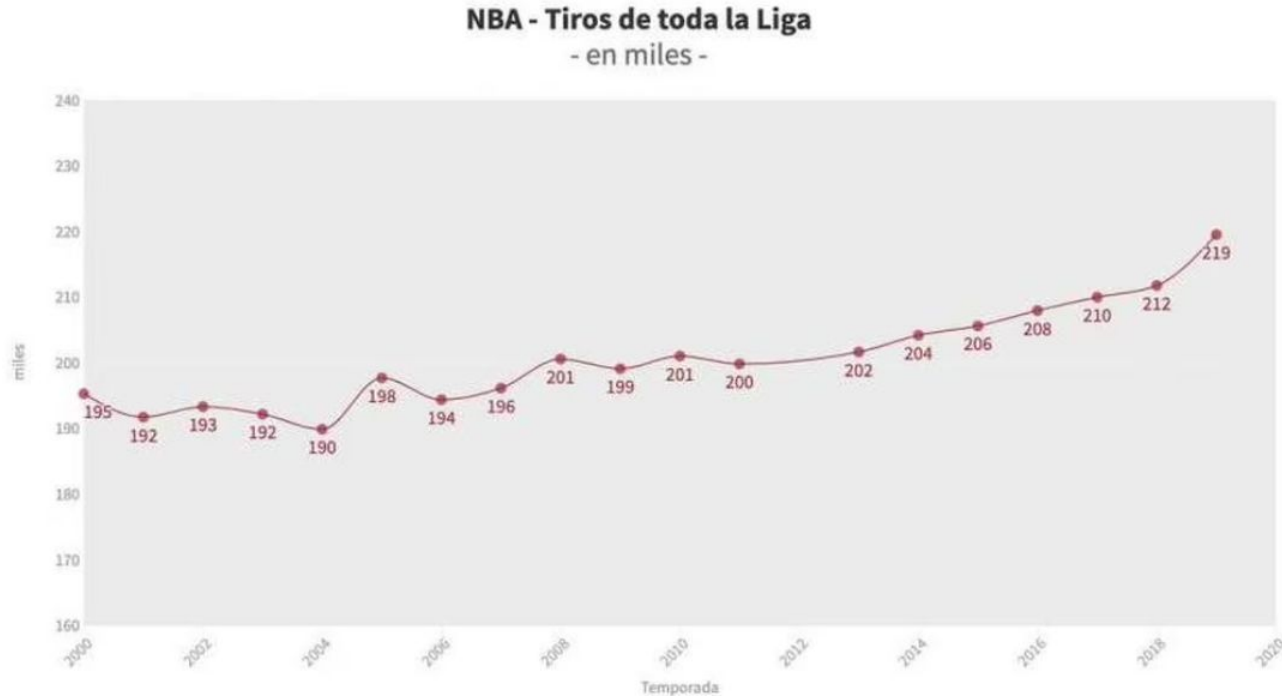
Año 2000



Año 2019



Caso introductorio 2: El juego de hoy en la NBA



En el 2019 hubo 25.000 tiros más por temporada que en el 2000.

Caso introductorio 2: El juego de hoy en la NBA

Dato:

En el año 2005, el equipo de los **Phoenix Sun** era el equipo más rápido. Tenía una filosofía de juego innovadora: jugar a 7 segundos o menos cada bola.

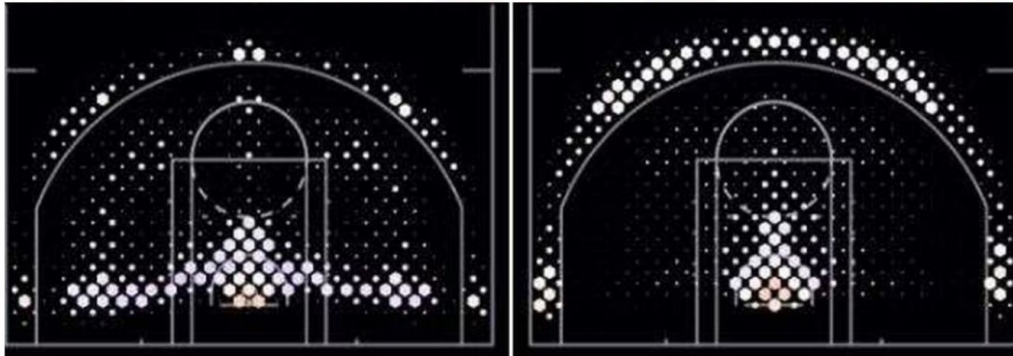
Insight 2019:

El equipo más lento de la NBA en 2019, **Charlotte Hornets**, jugaba más rápido que los **Phoenix Sun** en el 2005.

Caso introductorio 2: El juego de hoy en la NBA

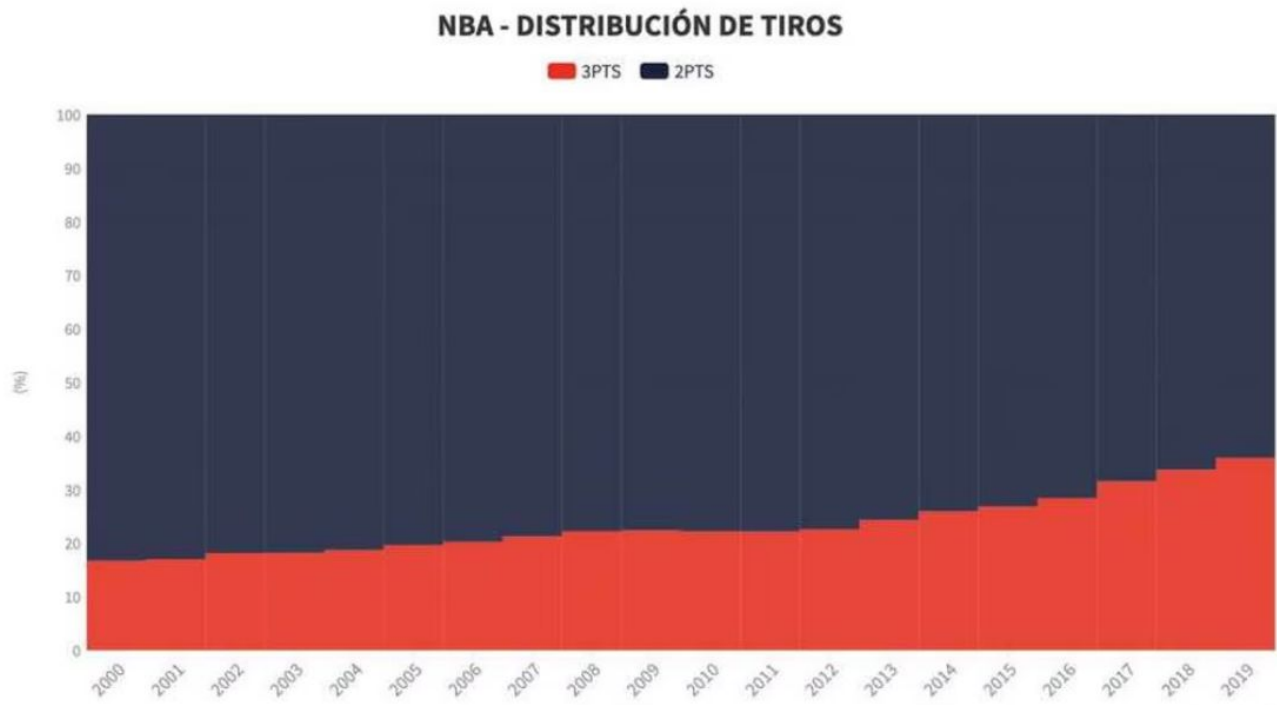
Año 2000

Año 2019



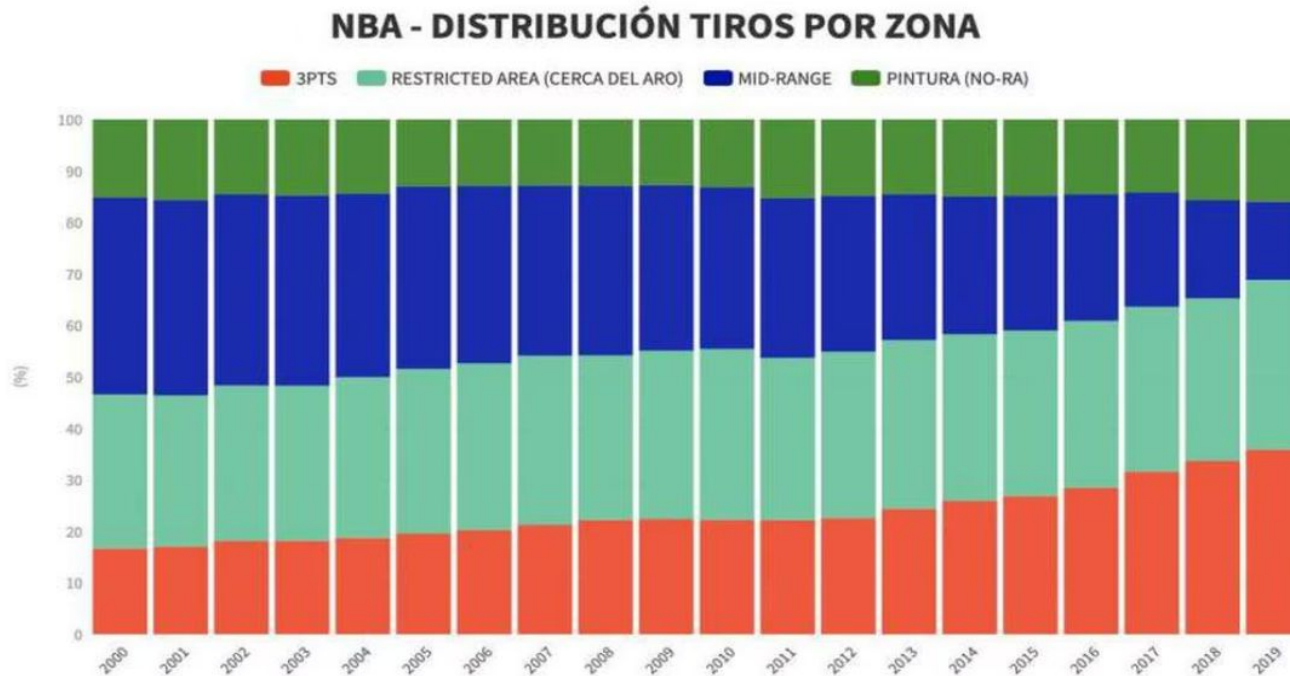
La clave está en el incremento de los triples como porcentaje del total de tiros.

Caso introductorio 2: El juego de hoy en la NBA



En el 2019, el 36% del total de tiros es de 3 puntos, contra el 16% del año 2000

Caso introductorio 2: El juego de hoy en la NBA



Insight:

Cambiar los tiros de media distancia por tiros de tres puntos.

Caso introductorio 3: Checkout QR MP

¿Todos han pagado alguna vez con QR desde la app de MP?



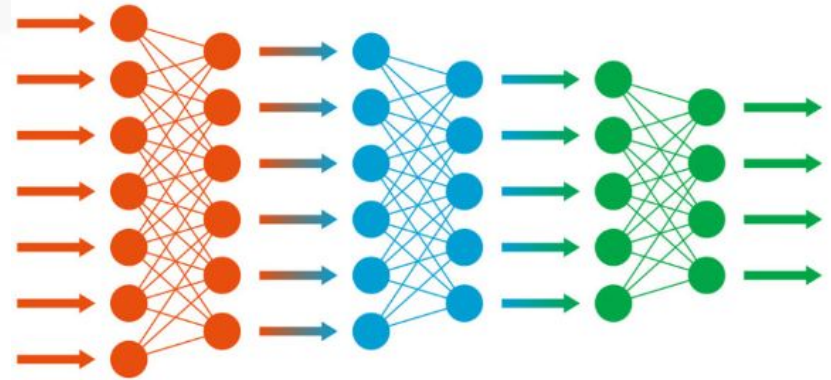
Caso introductorio 3: Checkout QR MP

- MP muestra un carrusel de medios de pago después del scan QR.
- El cliente hace swipe hasta el medio de pago preferido.



Caso introductorio 3: Checkout QR MP

- Modelo de machine learning que determina la probabilidad de uso de cada medio de pago.
- Los medios de pago se ordenan en un carrusel de acuerdo a la probabilidad de uso.



Caso introductorio 3: Checkout QR MP

Impacto:

- Mejora de la experiencia de usuario
- Disminución del número de swipes.
- Reducción del tiempo de checkout.
- Reducción de operaciones canceladas.



Minería de datos

Estos tres casos de uso muestran el potencial de la minería de datos:

- **Caso Walmart**, se descubrió un patrón oculto en los datos que incrementó las ventas de pañales y cervezas.
- **Caso NBA**, se comprendió una dinámica de juego más rápida, más precisa y con menos exigencia de talle a la hora de seleccionar un jugador.
- **Caso MP**, se mejoró la experiencia de usuario en la app de una manera personalizada.

Tarea #1: Casos de uso

Buscar información acerca de otros casos de usos emblemáticos (al menos tres) y describir:

- Hipótesis planteadas
- Descubrimiento
- Explicación del Patrón
- Acciones Tomadas
- Lecciones Aprendidas

En la próxima sesión veremos...

Repaso de Python, Git y Github:

- Uso de python en la nube: google colaboratory, anaconda
- Tipos estructurados (uso y mutabilidad)
- Programación Orientada a Objetos
- Git y Github

Preguntas