



¡Les damos la bienvenida!

¿Comenzamos?

Certificados oficialmente por

■ R/GA

CODERHOUSE

Esta clase va a ser

- grabada



COMISIÓN N°####

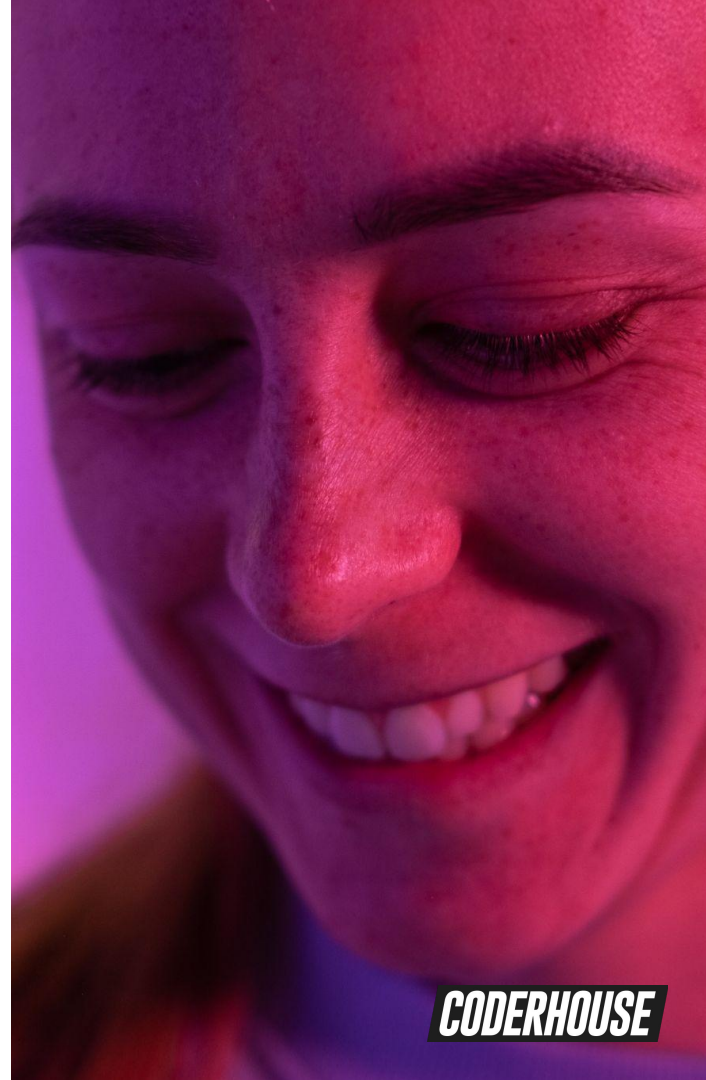
Presentación del equipo

- ✓ **Profesor/a responsable:** Juan Pérez
- ✓ **Coordinador/a:** Juan Pérez
- ✓ **Tutores y tutoras:**
 - Juan Pérez
 - ...
 - ...
 - ...
 - ...
 - ...
 - ...

Presentación de estudiantes

Por encuestas de Zoom

1. País
2. Conocimientos previos
3. ¿Por qué elegiste este curso?

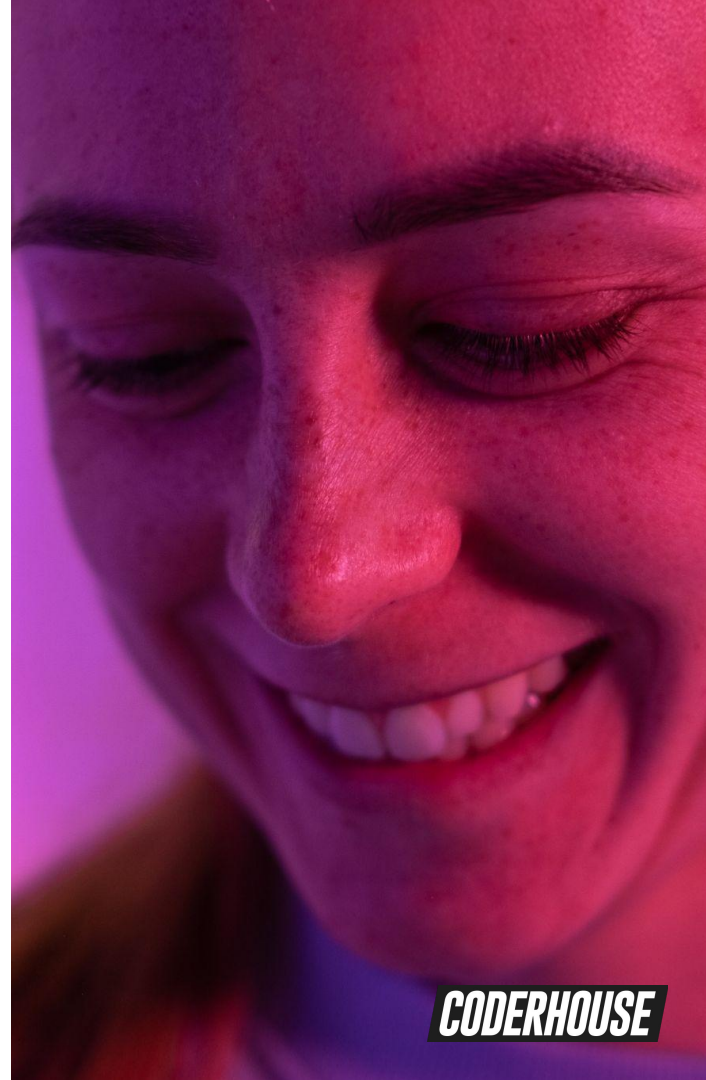


CODERHOUSE

Presentación de estudiantes

Por encuestas de Zoom

Conocemos un poco más respecto a nuestros
conocimientos y contexto.



CODERHOUSE

Unidad 01. FUNDAMENTOS DE LA CIENCIA DE DATOS

La Transformación Digital en la Industria 4.0

Certificados oficialmente por

■ R/GA

CODERHOUSE

Objetivos de la clase

- **Identificar** los componentes importantes de una estrategia de Data & Analytics.
- **Comprender** el rol de los datos en la organización desde una perspectiva de Transformación Digital y en la Industria 4.0
- **Facilitar** la identificación de oportunidades del uso de los datos para la transformación digital y la estrategia de negocios.
- **Clasificar** las principales herramientas para un Científico de Datos y sus características.



La **transformación digital** es el **fomento de evolución y nuevos modelos de negocio**, incorporando la **digitalización de archivos** e **incorporando lo digital a todas las áreas de negocio**.

– Harvard Business Review, 2021

Transformación Digital

La transformación digital es el proceso mediante el cual se realizan cambios integrales en la estrategia, modelos operativos, personas, cultura y procesos.

Problemas nuevos, exigen soluciones innovadoras 🚀

Deseamos ver mejoras dramáticas en el desempeño y cambiar las rutas para lograr el éxito.

Las amenazas para las organizaciones hoy en día son más resistentes y más robustas.

Como muestra el siguiente cuadro, existe un movimiento constante en el top de empresas líderes...

Ciclo de vida de un proyecto de Ciencia de Datos

Ciclo de vida de un proyecto de Ciencia de datos



Momento 1: Definir el objetivo



Momento 2: Recolección de la data



Momento 3: Preparar la data



Momento 4: Elección del Algoritmo

Ciclo de vida de un proyecto de Ciencia de datos



Momento 5: Entrenar el modelo



Momento 6: Validación del modelo



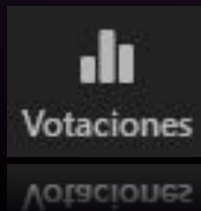
Momento 7: Deployment del modelo



¿Qué conceptos han escuchado antes?

Por encuestas de Zoom:

- ✓ Exploratory Data Analysis
- ✓ Data Wrangling
- ✓ Minería de datos
- ✓ Limpieza de datos
- ✓ SQL
- ✓ Estadística inferencial
- ✓ Machine Learning
- ✓ Deep Learning
- ✓ Inteligencia artificial





Ejemplo en vivo

¡Vamos a trabajar con datos! Miremos en los recursos complementarios “Unidad_1.ipynb” cómo crear un gráfico igual a este.

Polls

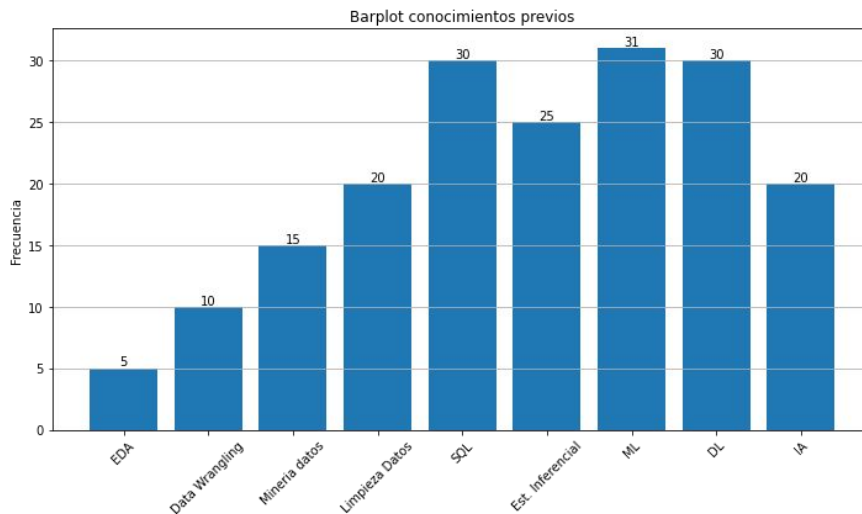
Encuesta conocimientos previos

Encuesta conocimientos previos

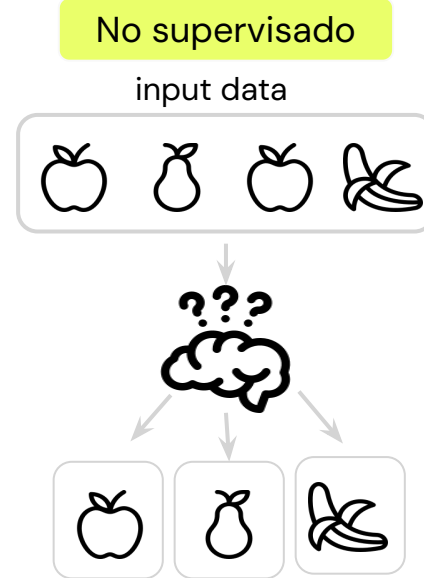
1. ¿Qué conceptos han escuchado antes? (Multiple Choice) *

- ☐ Exploratory Data Analysis
- ☐ Data Wrangling
- ☐ Minería de datos
- ☐ Limpieza de datos
- ☐ SQL
- ☐ Estadística inferencial
- ☐ Machine Learning
- ☐ Deep Learning
- ☐ Inteligencia Artificial

[Edit Poll](#) [Launch](#)



Aprendizaje supervisado y no supervisado





Para pensar

Supongamos que una empresa lanza una campaña de marketing para encontrar los diferentes segmentos para un nuevo producto.

¿Que tipo de aprendizaje deberíamos utilizar en este caso? ¿Por qué?

Contesta en el chat de Zoom



Optimizando el stock para una PYME

Ayudamos a optimizar el stock de nuestra heladería de barrio

Duración: **10 minutos**



ACTIVIDAD EN CLASE

Optimizando el stock para una PYME

Consigna: En la heladería de Pedro se lleva mucho tiempo trabajando sin ningún tipo de estrategia enfocada al uso de los datos como oportunidad de mejora y manejo de stocks. Recientemente, debido a la crisis sanitaria, el dueño ha cambiado su perspectiva y piensa que su empresa debería hacer un mejor uso de los datos históricos recolectados.

- ✓ ¿Cómo piensan que esta información puede ayudar a tener un mejor control del stock de la heladería?
- ✓ ¿Qué nivel de madurez tendría esta empresa según el modelo Data Management Maturity Model?

Bases de Datos

Ejemplo de consulta SQL

```
SELECT p.FirstName, p.LastName, a.City, cd.Detail
FROM Person p
JOIN ContactDetail cd ON cd.PersonId = p.Id
JOIN ContactDetailType cdt ON cdt.Id = cd.TypeId
JOIN Address a ON a.PersonId = p.Id
```

Ejemplo de consulta no SQL

```
{  
  "Id": "1",  
  "firstName": "Thomas",  
  "lastName": "Andersen",  
  "addresses": [  
    {  
      "line1": "100 Some Street",  
      "line2": "Unit 1",  
      "city": "Seattle",  
      "state": "WA",  
      "zip": "98012"  
    }  
  ],  
  "contactDetails": [  
    { "email": "thomas@anderson.com" },  
    { "phone": "*1 555 555-5555", "extension": 5555 },  
  ]  
}
```



Ejemplo en vivo

Realizaremos una consulta SQL en vivo para extraer información sobre los empleados de la tabla empleados.

Tiempo: 15 min

De los recursos complementarios utilizaremos el archivo “ventas_tienda”.



Para pensar

¿Qué casos de implementación en la industria conoces o has escuchado hablar, tanto de SQL como de NoSQL en empresas?

Escribe en el chat de Zoom

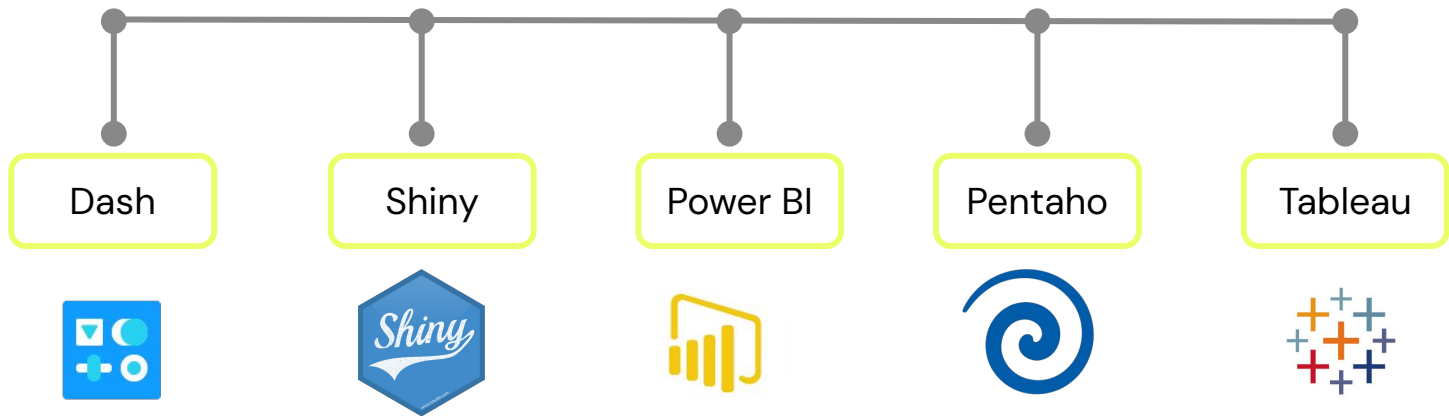


Break

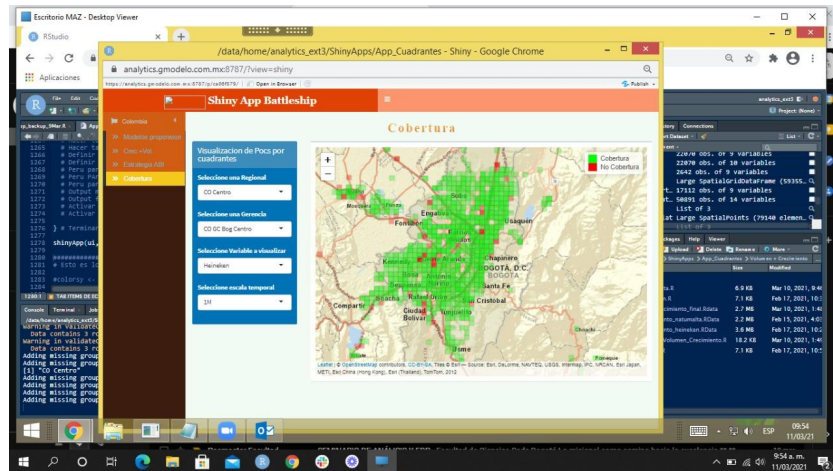
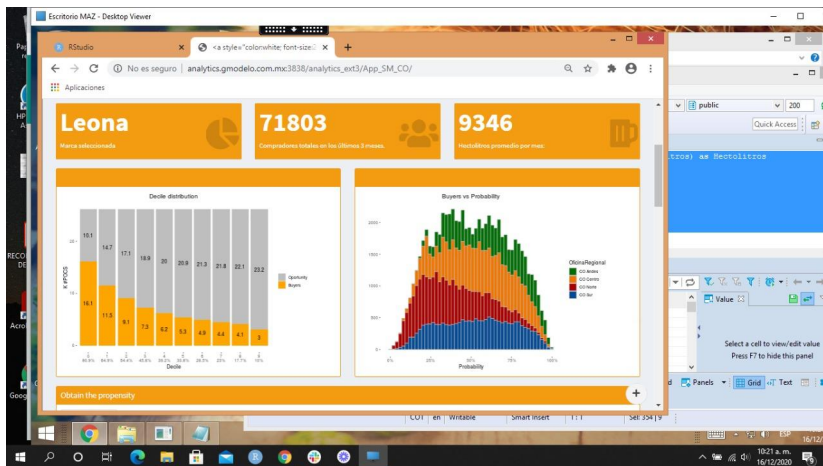
¡10 minutos y volvemos!

Visualización de datos

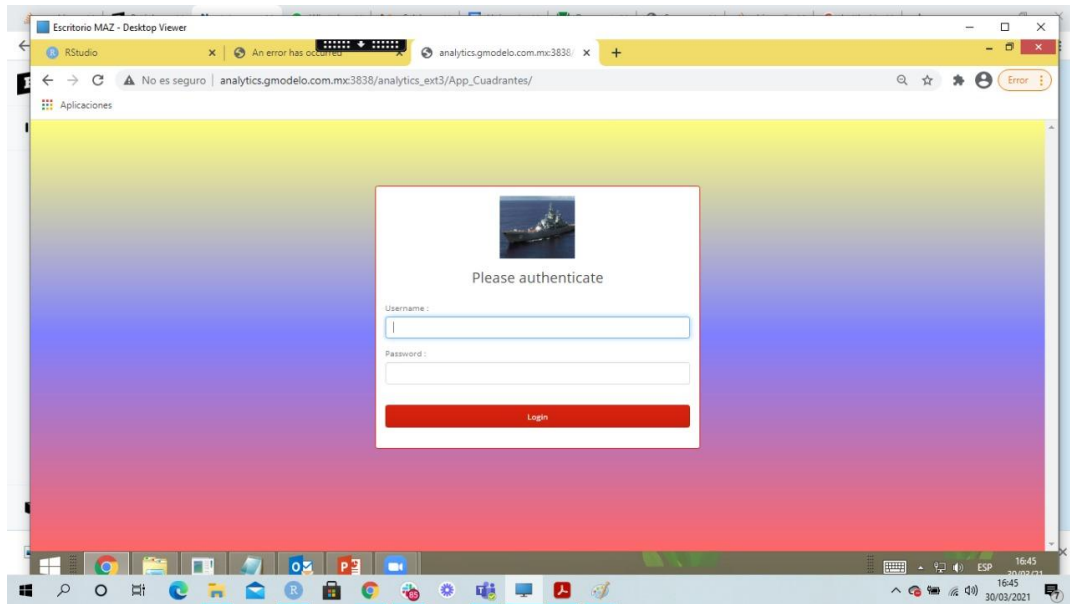
Herramientas de visualización de datos



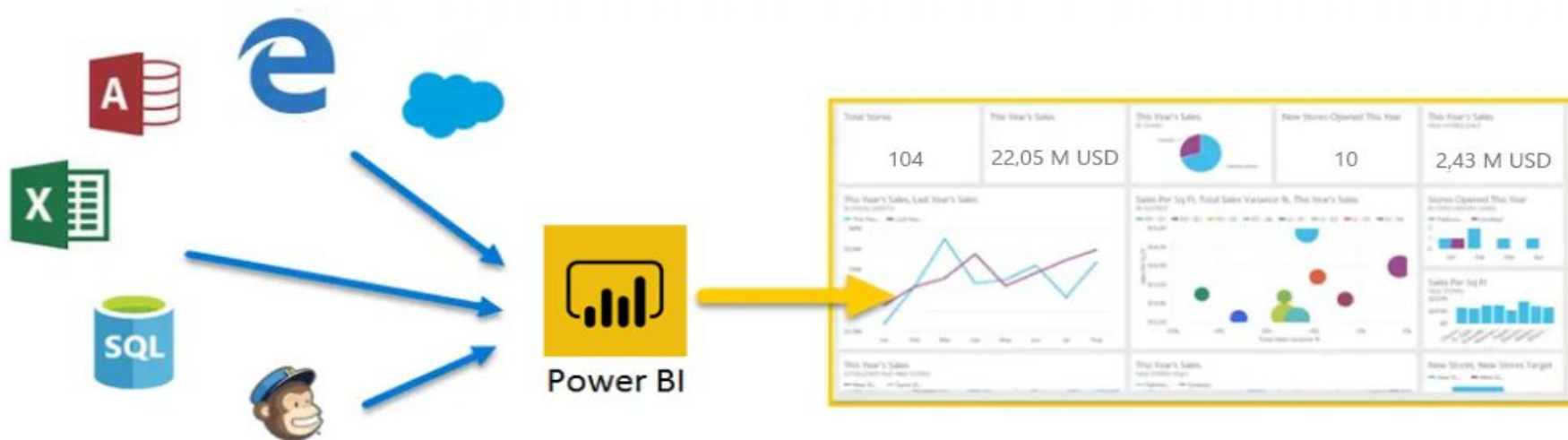
Ejemplos reales



Ejemplos reales



Ejemplo





Ejemplo en vivo

Vamos a ver cómo crear una cuenta en Github y el paso a paso para subir allí información.

Duración: **5 minutos**

Sistemas in-house

Ventajas y desventajas In-House

Ventajas	Desventajas
Seguridad. Manejo de datos sensibles sin terceros	Requiere de grandes inversiones iniciales para infraestructura y hardware
Sin tantos gastos fijos (pago a terceros por servicio)	Gastos variables por consumo y espacio además de mantenimiento
Posibilidad de autogestión dentro de la compañía.	Tiene una carga máxima (capacidad de procesamiento)
No requiere de conexión a Internet (no es limitante velocidad de conexión)	Requiere de personal capacitado para resolver problemas
Hay control sobre la infraestructura (control de acuerdo a las necesidades)	Pierde flexibilidad (acceso remoto) además de ser susceptible a pérdida de datos (daños estructura)

Cloud Computing



“El Cloud Computing, permite la aceleración y gestión de procesos computacionales, haciéndolos más eficientes”

- Observatorio Nacional de Telecomunicaciones y de la SI

¡Atención!

Recuerda instalar Python con Anaconda para la próxima clase.



Ver tutoriales en carpeta de clase

¿Preguntas?

Resumen de la clase hoy

- ✓ Cuarta Revolución Industrial
- ✓ El Ambiente de la Industria 4.0
- ✓ Transformación Digital
- ✓ Ciclo de vida de un proyecto de ciencia de datos
- ✓ Bases de Datos Relacionales y No Relacionales
- ✓ Lenguajes de Data Science
- ✓ Visualización de Datos
- ✓ Herramientas Complementarias.

Muchas gracias.

Opina y valora
esta clase