

## Objetivo del Proyecto

Realizar en grupos una presentación de **máximo 10 minutos** que describa el análisis de un problema de negocio basado en datos, siguiendo las fases de la metodología CRISP-DM

Grupos de 3 personas (Los mismos grupos del proyecto 1)

## Desarrollo del proyecto

Cada equipo deberá:

1. **Seleccionar y descargar** un dataset relevante a un problema real de negocio.
2. **Aplicar** cada una de las fases de CRISP-DM (excepto Deployment) al caso elegido:
  - a. Comprensión del negocio
  - b. Comprensión de los datos
  - c. Preparación de los datos
  - d. Modelado
  - e. Evaluación
3. **Elaborar** una presentación clara y concisa que comunique todo el flujo de trabajo.
4. **Conclusiones** sobre hallazgos, limitaciones y siguientes pasos.

## Evaluación del Proyecto

Ítem	Peso	Detalle
1. Dataset seleccionado	25%	Relevancia, calidad y documentación del origen de datos.
2. Calidad de la presentación	50%	Claridad, coherencia, timing ( $\leq 10$ min), uso de visualizaciones, presentación personal.
3. Conclusiones	25%	Profundidad del análisis, recomendaciones y próximos pasos.

## Fuentes de datos

## 1. Repositorios generales de datos

- **Kaggle Datasets** Millones de datasets en múltiples áreas (salud, finanzas, marketing, etc.).  
<https://www.kaggle.com/datasets>
- **UCI Machine Learning Repository** Clásicos para aprendizaje automático y ejemplos bien documentados.  
<https://archive.ics.uci.edu/ml/index.php>
- **Google Dataset Search** Buscador que indexa miles de repositorios, académicos y comerciales.  
<https://datasetsearch.research.google.com/>

## 2. Portales de datos abiertos gubernamentales

- **Datos Abiertos Colombia** Miles de tablas sobre salud, transporte, educación, medio ambiente, etc.  
<https://www.datos.gov.co/>
- **Data.gov (Estados Unidos)** Extenso catálogo de agencias federales: economía, clima, demografía...  
<https://www.data.gov/>
- **European Union Open Data Portal** Información sociodemográfica y económica de la UE.  
<https://data.europa.eu/en>

## 3. Plataformas especializadas por dominio

- **World Health Organization (WHO) – Global Health Observatory** Estadísticas de salud global, mortalidad, vacunación, etc.  
<https://www.who.int/data/gho>
- **World Bank Open Data** Indicadores de desarrollo: PIB, pobreza, educación, cambio climático...  
<https://data.worldbank.org/>
- **OpenStreetMap (Geoespacial)** Datos de mapas libres (calles, puntos de interés, fronteras).  
<https://www.openstreetmap.org/>

## 4. APIs y fuentes “en vivo”

- **Twitter API / X API**
- Extraer tuits por palabra clave o usuario para análisis de sentimiento.  
<https://developer.twitter.com/>
- **OpenWeatherMap** Datos meteorológicos históricos y en tiempo real por ciudad.  
<https://openweathermap.org/api>