**Entrega 1**

**Formulación inicial del proyecto de Ciencia de Datos**

## 1. Contexto y Problema

Explica brevemente el entorno (empresa, investigación, industria) y cuál es el problema por resolver.

Gracias a la Ley de Derecho de Acceso a la Información Pública (Ley 27.275), cualquier persona puede consultar todos los movimientos y creaciones de las Sociedades en Argentina. Actualmente, no existen herramientas que nos permitan analizar las Sociedades Argentinas de manera eficiente. Al ser información no estructurada, se necesitan herramientas sofisticadas, que permitan entender el contexto de los negocios en Argentina.

## 2. Stakeholders

Lista los principales interesados (usuarios finales, áreas de negocio, tomadores de decisión).

Los principales interesados son:

* Analistas Políticos, Económicos y Sociológicos.
* Empresas de servicios en búsqueda de Leads.
* Fundadores y Emprendores analizando el mercado

## 3. Pregunta de Datos

Reformula el problema en términos de datos: ¿qué queremos medir, explicar o predecir?

* ¿Cómo podemos estructurar y normalizar los movimientos de las Sociedades Argentinas?
* ¿Qué tendencias están tomando las Sociedades en el pais?
* ¿Cuántas empresas han tenido que anunciarse en liquidación?
* ¿Cuántas sociedades existen por cada rubro?

## 4. Objetivo General

Enuncia de forma clara y medible qué busca el proyecto.

Construir un proceso que permita extraer, normalizar y analizar las Sociedades Argentinas.

## 5. Objetivos Específicos

Lista entre 3 y 5 objetivos concretos.

* Normalizar y estructurar todos los movimientos y creaciones de Sociedades en Argentina.
* Automatizar y definir un pipeline de ETL robusto que permita replicar el dataset para cualquier investigador.
* Investigar y analizar las tendencias de las Sociedades Argentinas.

## 6. Alcance

Define qué datos, periodo temporal y variables incluirás. Menciona exclusiones.

## 7. Métricas de Éxito

Define cómo sabrás que el proyecto fue exitoso. Usa métricas cuantitativas.

* +10.000 Sociedades Argentinas normalizadas.
* +1.000 registros de alta fidelidad para utilizar en Finetuning.

## 8. Fuentes de Datos

Lista las fuentes, variables clave y observaciones sobre su calidad.

La fuente de datos es el Boletin Oficial. [https://www.boletinoficial.gob.ar](https://www.boletinoficial.gob.ar/)

## 9. Plan de Trabajo

Divide en etapas: Serán dependientes de la metodología que emplees. Por ejemplo: limpieza, EDA, modelado, evaluación, comunicación.

(Completar aquí)

## 10. Cronograma

(Completar aquí, formato de cuadro)

## 11. Riesgos y Limitaciones

Menciona problemas potenciales: calidad de datos, falta de recursos, privacidad.

(Completar aquí)

## 12. Entregables

Lista productos finales: dataset limpio, modelo, informe, dashboard, presentación.

(Completar aquí)

## 13. Comunicación

Define cómo y a quién presentarás resultados (formato y frecuencia).

(Completar aquí)

## 14. Recursos y Roles

Indica los recursos necesarios (humanos, tecnológicos, financieros) y los roles de cada integrante del equipo.  
Ejemplo: Data Scientist (modelado), Analista de Datos (limpieza y EDA), Ingeniero de Datos (ETL), Líder de Proyecto (coordinación), Recursos: servidores, software, presupuesto estimado