

Diplomado en Ciencia de Datos y Análisis

Avanzado

Unidad N°9 Proyecto Final Integrador

Proyecto Final Integrador



Objetivos

- **Definir** claramente el problema a resolver, estableciendo su relevancia para el negocio y los objetivos estratégicos asociados.
- **Seleccionar** y analizar un conjunto de datos adecuado, garantizando su calidad mediante procesos de limpieza, tratamiento de valores faltantes e ingeniería de variables.
- **Aplicar** una metodología estructurada (**CRISP-DM** u otra equivalente) para guiar el desarrollo del proyecto de forma coherente y reproducible.
- **Identificar riesgos y limitaciones** del proyecto, incluyendo consideraciones éticas y de privacidad de los datos utilizados.



Consigna

Se solicita a los equipos **definir el tema de su proyecto final**. El proyecto deberá centrarse en un problema de una empresa real o hipotética, o en un tema relevante que el equipo considere pertinente.

El proyecto final integrará los contenidos técnicos y de negocio, aplicando los conocimientos adquiridos a lo largo de la diplomatura.

Cada equipo deberá demostrar su trabajo mediante la elaboración y entrega de **tres recursos**:

1. **Informe Final escrito**, según normativa detallada.
2. **Código Python** en colab y enlace de repositorio público (GitHub).
3. **Video de 3 minutos** de presentación de resultados con PPT o similar.

Presentados de manera completa y profesional, mostrando claramente los resultados obtenidos y en lenguaje claro para directivos no técnicos.

A continuación, se detallan las instrucciones:

Pasos a seguir

1. Consultar la conformación del equipo

Revisar en Moodle el equipo al que pertenecen.

2. Definir el tema del proyecto

Discutir con su equipo la elección del tema y seleccionar uno.

Pueden elegir uno de los temas sugeridos o proponer uno propio, siempre que:

- Sea un problema relevante para un negocio real o hipotético.

- Disponga de datos suficientes para un análisis completo de ciencia de datos.

3. Elección del tema

Revisar los temas sugeridos (opcionales)

- **Pronóstico de ventas:** para tarjetas de crédito, telecomunicaciones o e-commerce.
- **Predicción de abandono (churn):** en RRHH o empresas de tecnología con alta rotación.
- **Segmentación e hiper-personalización de clientes:** mediante clustering y sistemas de recomendación.
- **Análisis de tiendas de alto desempeño:** identificación de factores que explican diferencias en facturación.
- **Detección de fraude:** en tarjetas de crédito, programas de fidelización, compras online o préstamos personales.
- **Clasificación de reclamos:** en e-commerce, logística, seguros o atención al cliente.
- **Predicción de mora:** en préstamos personales, fintechs, cooperativas o empresas de microcrédito.
- **Predicción de stock crítico:** identificación de productos que podrían agotarse por picos de demanda o errores en la reposición.
- **Influencer marketing analytics:** a partir de reseñas, redes sociales o encuestas de satisfacción.

- **Optimización de precios dinámicos:** detectar oportunidades para subir o bajar precios y maximizar ingresos.
- **Análisis del desempeño académico.** predicción de deserción o rendimiento en instituciones educativas.
- **Detección de anomalías en sensores o IoT:** especialmente en entornos agrícolas e industriales

Se puede elegir un tema diferente al de la lista, la lista es una base de ejemplos.

Una vez definida la temática del proyecto, corresponde avanzar hacia la etapa de pre-entrega, en la cual se evaluará la solidez del enfoque inicial.

4. Pre-entrega:

Objetivo de la pre-entrega:

- Comprobar, con evidencia mínima, que el equipo definió correctamente el problema y su KPI de negocio.
- Verificó la viabilidad de datos (acceso, calidad, licencias).
- Estableció un baseline reproducible y un plan de experimentación con métricas y validación.
- Identificó riesgos/limitaciones y un plan de trabajo para llegar al resultado final.

La pre-entrega no exige resultados “óptimos” ni storytelling pulido: prioriza solidez del enfoque, trazabilidad y factibilidad.

Se le dará al alumno una devolución de la preentrega para hacer sus ajustes de cara a la entrega final.

Estructura sugerida para la pre-entrega: Borrador breve (3–4 páginas, PDF)

1. **Resumen ejecutivo ($\leq \frac{1}{2}$ página):** problema, público interno/externo, KPI(s) de impacto esperado.
2. **Definición del problema y caso de negocio:** objetivo medible; hipótesis; restricciones; valor esperado
3. **Relevamiento de Fuentes de Datos a utilizar:** fuentes, tamaño, diccionario mínimo (top 10 variables), muestra/tablas de nulos, outliers, data quality issues, decisiones de limpieza ya tomadas y pendientes.

4. Plan de modelado y validación: métricas a reportar, esquema de CV/holdout, umbrales de éxito y benchmark de negocio.

Una vez realizado el borrador de la pre-entrega, las correcciones realizadas por el tutor servirán de base para la elaboración de su informe final y el resto del proyecto.

Respetar las longitudes máximas de cada etapa del trabajo.

5. Elaboración del Informe Final

Se deberá elaborar un informe final tomando en cuenta la siguiente estructura:

Estructura mínima Recomendada (10 páginas)

- **Portada (1 página):** Incluir el título del proyecto, nombre del equipo e integrantes, fecha y diplomatura o curso.
- **Resumen Ejecutivo (1 página):** Describir el problema a resolver, la solución propuesta, el impacto esperado en el negocio y los resultados clave, incluyendo métricas principales.

- **Definición del Problema y Relevancia:** Presentar un contexto breve y la motivación del proyecto, definir un objetivo concreto y medible, y explicar la importancia estratégica para el negocio.
- **Datos y Metodología:** Detallar la fuente y características de los datos, incluyendo tamaño, variables clave y calidad, incluir conclusiones de informe EDA; describir los procesos aplicados como limpieza, tratamiento de valores faltantes e ingeniería de variables; y explicar la metodología seguida, como CRISP-DM u otra similar.
- **Modelado y Evaluación:** Indicar los algoritmos probados y los criterios de selección, mostrar las métricas obtenidas (precisión, recall, F1-score, AUC, entre otras), y realizar una comparativa de modelos con justificación de la elección final.
- **Impacto en el Negocio:** Exponer los beneficios esperados, incluyendo ahorros, mejoras o automatización; identificar riesgos y limitaciones; y analizar la escalabilidad y los pasos futuros.
- **Conclusiones y Recomendaciones:** Resumir los hallazgos principales y proponer los siguientes pasos sugeridos.
- **Anexos opcionales:** Incluir gráficos adicionales, ejemplos de código, tablas extensas o diagramas que complementen la información del informe.
- **Código Python:** Proporcionar los scripts o notebooks alojados en un repositorio público (GitHub), incluyendo instrucciones en un archivo README para su correcta ejecución.

6. Video explicativo

Presentación grabada de máximo 3 minutos distribuidos el problema, enfoque metodológico, resultados e impacto esperado.

Evitar la lectura literal de texto y priorizar el storytelling con cifras clave. Se valorará la capacidad del grupo de científicos de datos para poner foco en lo relevante.

Requisitos técnicos y metodológicos

- Seguir la metodología CRISP-DM (o equivalente).
- Evaluar al menos dos algoritmos y justificar la selección final.
- Conjuntos de datos de menos de 3 años solo se usarán para series de tiempo si se explica la limitación; para otros casos se sugiere XGBoost, Random Forest u otros.
- El código debe poder ejecutarse localmente (no se exige despliegue en producción).
- Librerías recomendadas: pandas, scikit-learn, XGBoost, matplotlib/Plotly, y para clustering scikit-learn + umap-learn.

Recomendaciones de trabajo durante el desarrollo del proyecto

Trabajar con revisiones semanales internas, manteniendo registros detallados de decisiones (data logs). Redactar el **Informe Final del Proyecto** de forma **iterativa**: comenzar por definir el índice, incorporar los resultados y desarrollar la introducción. Solicitar retroalimentación temprana durante un período de pre-entrega destinado a revisiones, para ajustar el enfoque antes de la entrega final. Recordar que la calidad de la narrativa es tan importante como la precisión técnica; asegurar claridad y coherencia en la presentación del proyecto para facilitar la comprensión del valor del trabajo realizado y su impacto en el negocio.

Formato de presentación

Presentación Final del Proyecto: PDF generado a partir de una presentación (PowerPoint, **Canva** u otra herramienta similar), con un **máximo de 12 páginas**, sin incluir los apéndices ni el código Python.

Formato y presentación:

- **Fuente:** Sans Serif (Arial, Calibri o Helvetica), **12 pt** para el cuerpo y **14–16 pt** para los títulos.
- **Interlineado:** 1,5 en todo el documento.
- **Longitud Máxima:** 12 páginas.
- **Márgenes:** 2 cm
- **Portada:** debe incluir título, nombres de los autores y fecha
- **Video explicativo:** El **video explicativo** debe entregarse mediante un enlace (YouTube), con una **duración máxima de 3 minutos**, estructurado de la siguiente manera:
 - problema y relevancia;
 - metodología (datos, técnicas y herramientas);
 - resultados clave e impacto.

Se permite un máximo de **6 diapositivas o una demo visual breve**, y el video debe subirse al campus virtual.

Fecha límite de entrega:

Pre-entrega: 08 de septiembre de 2025

Entregas final: 06 de Octubre de 2025

No es posible para el cuerpo docente hacer extensiones sobre las fechas límites, en caso de necesitar una extensión hay que solicitarlo al centro de E-Learning.

Criterios de evaluación

La evaluación tiene un valor global de **10 puntos**, distribuyéndose de la siguiente forma:

- **Claridad en la definición del problema y alineación con el negocio (20%):** Evalúa qué tan bien se ha identificado y cuantificado el problema, su relación con los objetivos estratégicos y cómo se justifican los valores esperados y las restricciones del negocio.
- **Rigor en la preparación de datos, seguimiento de la metodología y justificación técnica (20%):** Se centra en la calidad del pipeline de datos, la limpieza, la ingeniería de variables y la documentación del proceso, así como en la adherencia a metodologías como CRISP-DM.
- **Selección y defensa del modelo (15%):** Considera la adecuación del modelo elegido, la comparación con alternativas, la claridad en la explicación de la elección y la justificación técnica basada en métricas y objetivos.
- **Calidad de las métricas (15%):** Evalúa si las métricas utilizadas son apropiadas para el tipo de problema, correctamente calculadas y comparadas, y si permiten validar los resultados de manera robusta.
- **Interpretación de resultados y cuantificación de valor (15%):** Analiza la capacidad de traducir los resultados técnicos en insights accionables, su relevancia para el negocio y la priorización de riesgos y oportunidades.
- **Consideraciones éticas y de privacidad (5%):** Mide la identificación y manejo de aspectos éticos y normativos, incluyendo privacidad de datos, riesgos potenciales y mitigaciones implementadas.
- **Impacto y claridad de la presentación (10%):** Evalúa la efectividad de la comunicación del proyecto, la claridad visual y narrativa, la estructura lógica y la capacidad de transmitir resultados e implicancias de negocio de forma comprensible y atractiva.

La nota final será grupal, reservándose la coordinación el derecho de ajustar individualmente en casos de evidente disparidad de esfuerzo.

