

IMT 2200 – Introducción a Ciencia de Datos

Semestre 2025-2

Profesor: Rodrigo A. Carrasco

Rúbrica Repositorio en GitHub (Entrega Inicial y Actualización Final)

El repositorio en GitHub deberá ser un espacio organizado, claro y completo que permita a cualquier usuario reproducir los análisis realizados en el proyecto, explorar el código y los datos utilizados, y entender los procesos y resultados. La estructura y presentación del repositorio son fundamentales para garantizar la transparencia, la reproducibilidad y el impacto del trabajo.

La evaluación del repositorio será según los siguientes criterios:

1. Organización y Estructura (25%):
 - Carpetas y Archivos: El repositorio está organizado en carpetas y archivos claramente identificables (por ejemplo: data/, notebooks/, scripts/, docs/, results/).
 - Estructura Coherente: Los archivos están organizados lógicamente y siguen una jerarquía clara.
 - README.md: Incluye un archivo README.md que describe claramente el propósito del repositorio, el contenido y las instrucciones para replicar los resultados.
2. Documentación (25%):
 - Comentarios en el Código: El código está bien comentado, explicando las funciones, pasos y lógica de los scripts.
 - Guía de Uso: Se proporcionan instrucciones claras y detalladas sobre cómo instalar dependencias, correr el código y reproducir los resultados.
 - Documentación clara y completa: El README explica el problema, los pasos del proyecto y cómo reproducirlo. Se citan correctamente fuentes externas (artículos, código, IA) y se explicita el rol de cada integrante.
 - Referencias: Incluyen enlaces a fuentes externas, bibliotecas usadas o artículos relevantes que respaldan el trabajo. Indica claramente el uso de fuentes externas, incluyendo fragmentos de código adaptados, referencias conceptuales y herramientas de inteligencia artificial generativa, si corresponde. Se valorará especialmente la transparencia en la declaración de prompts y el contexto en que se usaron.
3. Calidad del Código (20%):
 - Buena Práctica: El código sigue estándares de buena práctica en programación (nombres descriptivos de variables, modularidad, claridad).
 - Reproducibilidad: Los scripts y notebooks están diseñados para que cualquier usuario pueda ejecutarlos con las instrucciones provistas.
4. Datos y ETL (15%):
 - Disponibilidad de Datos: Los datos están incluidos en el repositorio (o se proveen instrucciones claras para descargarlos), con documentación sobre su origen y formato.

- Procesos de ETL: Los scripts de ETL están documentados y claramente explicados, permitiendo replicar el preprocesamiento realizado sobre los datos.
5. Resultados y Visualizaciones (15%):
- Resultados: Los resultados obtenidos (tablas, gráficos, modelos) están organizados en carpetas específicas y vinculados al análisis presentado.
 - Visualizaciones Reproducibles: Los scripts para generar visualizaciones están disponibles y generan gráficos que son consistentes con los reportados en la página web y el video.

La rúbrica final busca garantizar que el repositorio no solo cumpla con los requisitos básicos, sino que sea un ejemplo de profesionalismo en el manejo de proyectos de ciencia de datos.

Rúbrica Página Web

Cada equipo deberá crear una página tipo GitHub Pages para comunicar y publicar los resultados de su proyecto. Esta página tiene como objetivo relatar en forma efectiva los principales objetivos y hallazgos del proyecto, incluyendo texto narrativo y figuras, en un lenguaje y formato adecuados para un público general. Los códigos y datos estarán enlazados a la página web.

La evaluación de la página será según los siguientes criterios:

1. Presentación (20%): Realiza una presentación efectiva, novedosa y atractiva, explicando claramente los detalles en forma sucinta y soportada por referencias y datos.
2. Contenido (50%): La página entrega toda la información requerida por el proyecto, incluyendo la descripción del problema, los datos usados, la metodología seleccionada, el proceso de ETL usado, los resultados obtenidos y la interpretación de estos. Reconoce el origen de ideas, visualizaciones o fragmentos adaptados de otras fuentes. Si se utilizó IA generativa para apoyar la escritura de partes no analíticas, debe estar indicado de forma explícita.
3. Visualización (10%): los datos, análisis y resultados se presentan en forma gráfica seleccionando los gráficos adecuados y con la información necesaria para transmitir la información correcta.
4. Discusión (20%): se presenta una sección de discusión donde se plantean las extensiones que pueden salir de este trabajo, los problemas éticos que podrían estar asociados y las posibles fuentes de error que podrían existir.

Rúbrica Video

Cada equipo realizará un video con la presentación de su trabajo con una duración de 10 minutos. Esta presentación debe enfocarse en las contribuciones y resultados del proyecto más que en los aspectos técnicos, y destacar los hallazgos y conclusiones del proyecto, o mensajes centrales que la audiencia debería recordar.

La evaluación de la presentación será según los siguientes criterios:

1. Presentación (30%): Realiza una presentación efectiva y clara, explicando claramente los detalles en forma eficiente usando las visualizaciones adecuadas.
2. Descripción del problema (10%): La pregunta se encuentra bien explicada, resaltando su importancia y efectos.
3. Metodología (20%): Se explica el proceso de ETL y posterior EDA completo, justificando por qué las herramientas seleccionadas son las adecuadas y lo que se hizo sobre los datos.
4. Resultados (20%): Se presenta en forma clara los resultados obtenidos, su interpretación es clara y las implicancias que tienen son coherentes. Se presentan temas éticos que aparecen del uso de las herramientas o análisis desarrollado.
5. Análisis Crítico (20%): Analizan e interpretan correctamente los resultados. Explican “¿Qué podría salir mal?” con su trabajo o su uso.