

Consignas del trabajo final de Ciencia de Datos para Economía y Negocios. FCE-UBA.

Docente: Nicolás Sidicaro

Introducción

El presente trabajo práctico tiene como objetivo que los estudiantes puedan consolidar de manera integral los conocimientos adquiridos durante la cursada y aplicarlos a un contexto específico de análisis de datos. Asimismo, se busca desarrollar las habilidades de comunicación de resultados a audiencias generales, una competencia fundamental en el ámbito profesional.

Este trabajo también constituye una oportunidad para que los estudiantes construyan un portfolio técnico que puedan presentar junto a su CV al momento de buscar oportunidades laborales en el campo de la ciencia de datos, analytics o áreas afines.

Conformación de equipos

Los equipos deberán estar conformados por **2 (dos) integrantes**.

Formato de entrega

La entrega consistirá en una **presentación (tipo PPT) en formato PDF** de la extensión que consideren apropiada para abordar completamente todos los puntos requeridos.

El objetivo principal es desarrollar habilidades de síntesis y comunicación efectiva para audiencias no técnicas. En este sentido, se valorará especialmente:

- **Evitar slides sobrecargadas de texto**
- **Utilizar formato de bullets** siempre que sea posible sin perder claridad conceptual
- **Combinar texto explicativo con gráficos** de manera equilibrada
- **Mantener un lenguaje claro y accesible** para audiencias generales

Estructura mínima requerida

Cada trabajo práctico deberá incluir, como mínimo, las siguientes secciones. Cada sección podrá ser subdividida a su vez en múltiples partes:

1. Introducción y objetivos
2. Descripción de los datos utilizados
3. Metodología aplicada

- a. En este punto se deberá justificar por qué se realizó determinado análisis, en función de los objetivos planteados en el punto 1
4. Análisis exploratorio de datos (EDA)
5. Resultados principales
6. Conclusiones
7. Limitaciones y trabajo futuro

Organización de archivos y códigos

Los equipos deberán organizar sus materiales siguiendo la siguiente estructura de carpetas:

- **/raw**: Archivos de datos en su formato original, sin procesar
- **/input**: Archivos de datos preprocesados y listos para el análisis
- **/output**: Resultados generados (modelos de ML, estadísticas descriptivas, gráficos, etc.)
- **/scripts**: Códigos utilizados en el proyecto

Importante: Los scripts deberán nombrarse indicando claramente el orden de ejecución mediante numeración al inicio del nombre. Ejemplos:

- 01_extraccion_datos.R
- 02_limpieza_de_datos.R
- 03_estadisticas_descriptivas.R
- 04_modelado_ml.R

Propuestas de trabajo

El trabajo presenta tres modalidades posibles: (i) analizar una base de datos propuesta por el docente; (ii) realizar el *scraping* de un sitio propuesto por el docente y el análisis de la información bajada; (iii) proponer un proyecto propio que sirva para sus trabajos y/o tesis de grado, se deberá proponer antes del 3 de junio para ser evaluado por el docente y aprobar su realización.

Opción 1

Se deberá analizar alguna de las siguientes bases de datos:

- a) Resultados de las pruebas Aprender 2023 ([link](#))
- b) Delitos en la Ciudad de Buenos Aires entre 2019-2023 ([link](#))
- c) Mapa Productivo Laboral de Argentina ([link](#))
- d) Viajes en subte ([link](#))
- e) Evolución del mercado inmobiliario en la Ciudad ([link](#))
- f) Flujo vehicular en la Ciudad ([link](#))
- g) Producto Bruto Geográfico (municipal) de la Provincia de Buenos Aires ([link](#))

- h) Datos de comercio exterior de las exportaciones de países de América Latina (CEPII-BACI) ([link](#)). Advertencia: la base es muy grande, quizás tengan que usar `data.table` para abrirla y filtrar los datos
- i) Precios de la última semana ([link](#))
- j) Otros datos de similar complejidad que pueden provenir de fuentes como OECD Data, World Bank Data, [Gapminder](#), Our World in Data
- k) Otros datos (si tienen dudas, consultar con docente)

Opción 1: versión a - Gráficos

Una vez seleccionada la base de datos, deberán realizar un análisis exploratorio de forma tal de comprender y poder transmitir las características de la base de datos, así como también las decisiones en cuanto a filtrado de datos que quieran realizar.

En esta opción deberán, luego del análisis exploratorio, enfocarse en contar una historia con gráficos llamativos, que no sean los gráficos clásicos, y que cumplan con los criterios estéticos vistos en clase y recomendados en los libros compartidos. Deberán justificar por qué se utilizó cada gráfico y las decisiones estéticas tomadas.

Opción 1: versión b - Tablero

Una vez seleccionada la base de datos, deberán realizar un análisis exploratorio de forma tal de comprender y poder transmitir las características de la base de datos, así como también las decisiones en cuanto a filtrado de datos que quieran realizar.

En esta opción deberán, luego del análisis exploratorio, enfocarse en contar una historia con gráficos en pocas diapositivas y como complemento hacer foco en realizar una visualización interactiva que pueda ser compartida al docente. Se deberá mencionar en la PPT por qué se escogió visualizar esos datos, cuál es el sentido y la historia que quisieran contar con los datos. Así como también justificar por qué se visualizaron con un tablero dinámico ¿qué suma hacerlo dinámico?

Opción 1: versión c - Machine Learning

Una vez seleccionada la base de datos, deberán realizar un análisis exploratorio de forma tal de comprender y poder transmitir las características de la base de datos, así como también las decisiones en cuanto a filtrado de datos que quieran realizar.

En esta opción deberán, luego del análisis exploratorio, enfocarse en aplicar al menos dos algoritmos de Machine Learning vistos en clase, ya sea para clasificar o para regresar. Deberán evaluar si el algoritmo logró el cometido buscado y justificar por qué lo utilizaron y los resultados obtenidos.

Opción 2: scraping y análisis

Se deberán realizar *scraping* de las siguientes categorías, con el objeto de comparar al menos dos páginas:

- a) Libros
 - i) [Cúspide](#)
 - ii) [Yenny](#)

- iii) [Buscalibre](#)
- b) Tecnología
 - i) [Gadnic](#)
 - ii) [Bidcom](#)
 - iii) [Provincia compras](#)
 - iv) [Fravega](#) (si eligen Fravega van a tener que hacer 3, ya que el código ya está compartido)
 - v) [Megatone](#)
- c) Alimentos
 - i) [La cooperativa obrera](#)
 - ii) [Jumbo](#)
- d) Celulares (dado que hay menos modelos, van a tener que comparar modelos específicos)
 - i) [Tienda Claro](#)
 - ii) [Personal](#)
 - iii) Fravega

Para cualquier opción de categorías que elijan deberán comparar al menos dos páginas, haciendo foco en la comparación de cada una de las categorías y -en algunos casos, como libros y celulares- de los productos específicos.

Se deberá detallar con claridad el proceso de *scraping*, la metodología utilizada y las dificultades encontradas. Asimismo, se deberá realizar una limpieza del texto para comparar productos, ya que no figuran con el mismo nombre en cada una de las páginas.

En el caso de los alimentos, se deberá tener cuidado en comparar peras con peras y no peras con manzanas. En otros productos, analizar las ofertas, la disponibilidad de cuotas sin interés como un diferencial de precios, entre otros.

Asimismo, los resultados deben estar claramente graficados, cumpliendo las normas de visualización vistas en clase. Otra opción es -al margen de la PPT que deberán entregar- hacer un tablero dinámico donde muestren los gráficos.

Se deberá mencionar en la PPT por qué se escogió visualizar esos datos, cuál es el sentido y la historia que quisieran contar con los datos. Así como también justificar por qué se visualizaron con un tablero dinámico ¿qué suma hacerlo dinámico?

Opción 3: trabajo libre

En caso de estar realizando la tesis o un trabajo puntual en tu ámbito laboral que pueda ser mezclado con el contenido de la materia, se puede hacer un trabajo específico para que se adecúe a las necesidades. Para eso, se deberá presentar una propuesta antes del 3 de junio para ser evaluada y confirmada por el docente.

Valoración adicional

Se otorgará **puntaje extra** a aquellos equipos que suban todo su proyecto a un **repositorio de GitHub**. Esta práctica no solo facilita la colaboración durante el desarrollo del trabajo,

sino que también les permitirá contar con un repositorio público con un análisis técnico, accesible y profesional para presentar ante eventuales propuestas de trabajo.

*Fecha de entrega: **tendrán tiempo para entregarlo hasta 04/07/2025**. Se puede entregar antes sin ningún inconveniente. Sin embargo, la entrega temprana no representa un “valor” extra, sino que representa una chance de que ustedes se ordenen más con las demás materias cursadas*

Modalidad de presentación: se deberá enviar por mail (nsidicaro.fce@gmail.com), tanto la presentación como las carpetas arriba mencionadas. Si se cargó todo en un Github se deberá también compartir esa carpeta de Github haciéndola pública.