

Práctica 2

Curso de Data Jurídica

ENERO 2026

Profesor: Enrique García Tejeda

Fecha de entrega: miércoles 28 de enero antes de las 11.59 pm

Requisitos generales para esta práctica

1. Identificación del creador del archivo
 - o Incluye tu nombre completo al inicio del script.
2. Código con comentarios
 - o Todo el código debe estar debidamente comentado para explicar qué hace cada sección.
3. Formato de entrega
 - o El archivo debe entregarse en un script con extensión .R.
 - o Nombra tu archivo de la siguiente forma:
ApellidoNombre_PracticaX.R (ejemplo: DavidCervantes_Practica1.R).
 - o Envía tu archivo al siguiente correo: david.cervantes@alumnos.cide.edu
4. Restricción de uso de herramientas de IA
 - o Para esta práctica no está permitido el uso de *large language models* (LLMs), como ChatGPT, Gemini, Claude u otros asistentes automáticos de programación, para generar el código o resolver directamente los ejercicios.
 - o El objetivo es que desarrolles las habilidades básicas de programación en R de forma autónoma y colaborativa.
 - o Se fomenta que consultes con tus compañeros de clase y que utilices manuales y documentación de R. Ejemplo recomendado: *Hands-On Programming with R*, Capítulo 2: "[The Very Basics](#)".
5. Envío
 - o Envía tu archivo al siguiente correo: david.cervantes@alumnos.cide.edu
 - o Título del correo: Curso Data Jurídica-Práctica 2

Objetivo de la práctica

Aplicar técnicas básicas de minería de texto y análisis de sentimientos en guiones de series de TV, usando R y los paquetes quanteda, quanteda.textplots, dplyr y textdata.

Para este trabajo sólo puedes elegir una de las siguientes series, que están incluidas en la base de datos de la carpeta “episodes”, que usamos en la sesión 2.

1. Better Call Saul
2. Chernobyl
3. Mindhunter

Ejercicios

1. Exploración inicial de la base de datos

- Carga el archivo .csv de la serie elegida.
- Explora la base con names(), head(), y View().
- Explica en un comentario:
 - o ¿Cuántas observaciones tiene el archivo?

- o ¿Cuántas variables tiene el archivo?
- o ¿Qué columnas trae el archivo?

2. Palabras clave y frecuencia

- Tokeniza el texto de la columna de descripciones.
- Haz al menos dos búsquedas con kwic() de palabras importantes (ejemplo: personajes o conceptos centrales de la serie).
- Calcula las palabras más frecuentes con topfeatures()

3. Limpieza y nube de palabras

- Limpia los tokens eliminando puntuación y stopwords en inglés.
- Genera un nuevo dfm y produce una nube de palabras con textplot_wordcloud().
- Explica: ¿qué emociones o temas aparecen reflejados en las palabras más visibles?

4. Análisis de sentimientos con diccionario NRC

- Construye un diccionario en español a partir de lexicon_nrc() con 10 categorías (enojo, alegría, miedo, etc.).
- Calcula los valores de positivo, negativo y total para cada episodio.
- Genera un gráfico de barras con la suma de positivos y negativos por episodio. Puedes usar los colores que tú quieras en los gráficos.
- Explica:
 - o ¿Predominan los sentimientos positivos o negativos en la serie?
 - o ¿Hay variaciones según la temporada?

Interpretación final

Escribe en máximo 10 líneas tu interpretación de los resultados:

- Busca en un LLM o en algún buscador común la trama de la serie si no la has visto
- ¿Crees que el análisis de sentimientos refleja bien la trama de la serie?