**Goodreads: Análisis de Libros y de Reviews**

***Para poder ejecutar el código, hace falta instalar wordcloud***

**pip install wordcloud *(para poder generar la nube de palabras) y vaderSentinent* pip install vaderSentiment *para poder hacer el análisis de opiniones.***

***A parte serán necesarios BeautifulSoup, Matplotlib, Plotly, Pandas y Streamlit.***

**1. Proceso para la extracción, tratamiento y preparación de los datos:**

* *Extracción:* Usando Python (3.12) con BeautifulSoup para obtener los datos directamente desde la página de Goodreads https://www.goodreads.com/book/most\_read?category=all&country=X&duration=X. A través de un dashboard en Streamlit, se selecciona un país y un período (últimos 12 meses, este mes o esta semana) y el sistema recopila automáticamente los 50 libros que aparecen en el ranking, junto con datos para cada libro (accediendo a través del href) como título, autor, calificación promedio, géneros y las primeras 30 reseñas de los lectores. Todo se guarda en un archivo CSV, algunos como valores únicos y otros, como las reviews o los géneros, en listas.
* *Limpieza:* Una vez tengo los datos, filtro los valores incompletos y preparo los textos de las reseñas. Por ejemplo, para una visualización de las palabras más comunes quito palabras comunes (en español y en inglés) y que no aportan información específica de cada libro como "de", "en" o "very" mediante una lista de “stopwords” y organizo los datos para que sean más fáciles de analizar.
* *Procesamiento:* Hago varios análisis en relación con los libros y con las reviews. Por ejemplo, un análisis de los géneros más populares, de los años de publicación más comunes en los rankings, el número de páginas que suele tener cada género y por cada libro individual un análisis de opiniones para clasificar las reseñas como muy positivas, positivas, neutrales, negativas y muy negativas, así como la opinión general del libro cogiendo la clasificación mayoritaria. También destaco las palabras más frecuentes, lo que ayuda a identificar temas recurrentes en los comentarios.

**2. Visualizaciones**

Para que los resultados fueran claros y fáciles de interpretar, hay varios tipos de visualizaciones:

* *Gráfico de barras de los 10 géneros más populares:*  
  Muestra los géneros que más aparecen en el ranking de libros. Es útil para identificar tendencias literarias y los tipos de libros que más atraen a los lectores.
* *Gráfico polar de los géneros más populares:*  
  Proporciona una perspectiva visual diferente para destacar los géneros más relevantes.
* *Gráfico de barras apiladas por año de publicación y género:*  
  Permite analizar cómo han evolucionado los géneros más populares a lo largo del tiempo. Sirve para observar cambios en las preferencias literarias de los lectores.
* *Gráfico de barras de libros publicados por año:*  
  Enseña la cantidad de publicaciones de los libros más populares por año, mostrando de que años suelen ser y si hay libros muy antiguos que mantienen su popularidad o que se han puesto de moda.
* *Gráfico de dispersión entre número de páginas y géneros:*  
  Explora la relación entre la extensión de los libros y su género (cogiendo el primero que se puede ver en Goodreads). Es útil para detectar patrones, como si ciertos géneros tienden a tener libros más largos o cortos.
* *Gráfico de barras con las palabras más mencionadas en las reseñas:*  
  Muestra las 20 palabras que más se repiten en los comentarios de los lectores. Sirve para entender qué aspectos del libro llaman más la atención mediante palabras.
* *Nube de palabras (wordcloud):*  
  Las palabras más frecuentes se ven más grandes y resaltan visualmente. Es una forma rápida y llamativa de captar las ideas principales que mencionan los lectores.
* *Gráfico de barras sobre las opiniones:*  
  Muestra cómo se distribuyen las reseñas entre "muy positivo", "positivo", "neutral", "negativo" y "muy negativo" para cada libro. Sirve para ver de un vistazo si un libro tiene buena acogida o deja mucho que desear.
* *Resumen de la opinión general:*  
  Enseña el sentimiento predominante en las reseñas. Es útil para ver cómo se siente en general la gente hacia un libro en particular.

**3. Conclusiones / Insights:**

Analizando las visualizaciones de todo el mundo y todos los países disponibles en la selección para los últimos 12 meses (para observar una selección de tiempo estándar), se pueden observar varios *insights* interesantes:

* Los géneros más populares son siempre **fiction** y **audiobook** y casi siempre **romance**, en orden.
* Los libros que son populares ahora fueron publicados mayoritariamente en **2023 y 2024** para todos los países, y en ciertos países hay libros muy antiguos (de hace uno / dos siglos) que siguen siendo muy relevantes actualmente.
* Para algunas selecciones de región hay géneros que se mueven en rangos de números de páginas similares, por ejemplo, en todo el mundo, los de **thriller**, **romance** y **fiction**.
* Entre las palabras más usadas en las reviews de cada libro suelen estar el nombre o los nombres de los protagonistas.
* Las opiniones más recurrentes de las reviews más populares que se muestran para cada libro en Goodreads suelen ser de tono “**muy positivo**”.