**Tarea: Análisis de Texto de los Programas de Gobierno**

**Instrucciones:** Utilizando los Programas de gobierno 2025-2030 de los siguiente partidos políticos, realiza las siguientes actividades de análisis de texto. Asegúrate de justificar tus respuestas con el código y los resultados obtenidos y explicar cada decisión.

En este caso la pregunta que guía este análisis es conocer un poco más sobre los énfasis de cada partido, y las similitudes/diferencias entre cada propuesta.

Links:

* Frente Amplio: <https://www.frenteamplio.uy/programa/propuesta-de-bases-programaticas-aprobadas-por-el-plenario-para-construir-el-uruguay-del-futuro/>
* Partido Nacional: https://alvarodelgado.uy/pdf/Programa-Delgado.pdf
* Partido Colorado: <https://programapartidocolorado.uy/wp-content/uploads/2024/08/LIBRO-PROGRAMA-UNICO-DE-GOBIERNO-2025-PARTIDO-COLORADO.pdf>
* Partido Independiente: <https://partidoindependiente.uy/wp-content/uploads/2024/05/Programa.-pdf-1.pdf>
* Cabildo Abierto: <https://cabildoabierto.uy/wp-content/uploads/2024/08/Programa-CA-2025-2030.pdf>
* No encontré link a programas de Identidad Soberana, Unidad Popular, Partido Avanzar Republicano, Partido Verde Animalista, PERI. Si los encuentran y quieren añadir, siéntase libres!

1. **Abrir los archivos:**
   * Investiga cómo convertir los pdfs a texto para su procesamiento.
2. **Exploración Inicial de los Textos:**
   * a) Carga los textos de todos los partidos y realiza un análisis exploratorio inicial. ¿Cuál es el número total de palabras en cada manifiesto? ¿Hay alguna diferencia significativa en la longitud de los manifiestos?
   * b) Identifica y elimina las palabras vacías (stop words). Luego, muestra las 10 palabras más frecuentes en cada manifiesto. ¿Observas alguna palabra que sea particularmente destacada en un manifiesto en comparación con los otros?
   * c) Genera una nube de palabras para cada uno de los manifiestos, destacando las palabras más frecuentes después de eliminar las palabras vacías. ¿Qué términos parecen ser centrales para cada partido según sus nubes de palabras?
   * d) Basado en c) y d), ¿hay alguna palabra que haya que incluir a las stop words? Inclúyelas y realiza nuevamente los pasos b, c y d hasta estar satisfecho.
   * e) ¿Qué pasa con b) cuándo lo hacemos luego de la transformación tf-idf?
3. **Análisis de Similaridad:**
   * a) Calcula la similaridad de coseno entre los manifiestos de los tres partidos. ¿Qué dos manifiestos son los más similares? ¿Qué dos son los menos similares? Interpreta estos resultados en el contexto de las posiciones políticas de los partidos.
   * b) Basado en la similaridad, ¿podrías inferir algo sobre la alineación ideológica entre los partidos? ¿Qué partido parece estar más distante de los otros?
4. **Análisis de Frecuencia de Bigrams:**
   * a) Identifica los bigramas (pares de palabras consecutivas) más frecuentes en cada manifiesto. ¿Hay algún bigrama que aparezca con frecuencia en los tres manifiestos? ¿Qué bigramas son únicos o distintivos de cada partido?
5. **K-Means**
   * a) Genera dos clusters utilizando K-means y analiza cómo quedan alineados los partidos. ¿Tiene sentido?
   * b) ¿Cuáles son las palabras más frecuentes en cada cluster?

**Siempre justifica que decisiones tomas.**