En este directorio se encuentran los notebooks correspondientes a la primera sección del proyecto, el análisis de reportes de sustentabilidad en pdf.

El trabajo inicial es forzosamente manual, se requiere de un listado de empresas y de los enlaces a sus reportes. Para ello basta con buscar en Google la pagina oficial de la empresa en cuestión y explorar su sitio para encontrar el reporte. Los reportes se deben ubicar en la carpeta “pdf\_es” que está en el mismo directorio que ambos notebooks.

Tendiendo esto el primer paso es realizar la extracción y limpieza del texto, para ello ejecutar el notebook "text\_extract.ipynb".

En caso de querer agregar o eliminar empresas solo es necesario modificar las celdas donde se definen las variables "esg\_urls\_rows" y "pages" (celdas 2 y 5)

Para realizar el análisis de los reportes se ejecuta el notebook "topic\_analysis.ipynb"

Es necesario ajustar la variable "company\_names" en la segunda celda de código para que incluya los mismos nombres de empresa utilizados en el notebook anterior (también calza con el nombre que se les da a los archivos .feather generados).

Las instrucciones específicas de ejecución se encuentran dentro de los notebooks.