Análisis Resultado del Script

Analisis Detallado del Resultado del Script y los Datos del Excel:  
  
### Ejemplo de Datos Obtenidos (10 Archivos PDF Aleatorios):  
  
| Nombre del Archivo | Peso (KB) | Tipo |  
|---------------------------------------|-----------|--------|  
| pdf/Ver\_Documento\_38255.pdf | 1244.36 | Texto |  
| pdf/109969\_11351.pdf | 382.58 | Texto |  
| pdf/81055\_67155.pdf | 652.72 | Texto |  
| pdf/Ver\_Documento\_37979.pdf | 3008.23 | Texto |  
| pdf/81169\_67173.pdf | 2723.27 | Texto |  
| pdf/Ver\_documento\_36849.pdf | 3385.24 | Texto |  
| pdf/Ver\_documento\_21936.pdf | 246.53 | Texto |  
| pdf/Ver\_Documento\_37885.pdf | 1382.05 | Texto |  
| pdf/Ver\_Documento\_38677.pdf | 14326.80 | Imagen |  
| pdf/Ver\_Documento\_38105.pdf | 1698.26 | Texto |  
  
  
### Análisis Estadístico Detallado:  
  
1. Distribución de Tipos de Archivos PDF:  
 - Archivos de texto (tipo 1): 848  
 - Archivos de imagen (tipo 0): 113  
  
2. Estadísticas Básicas del Tamaño de los Archivos PDF (en KB):  
 - Tamaño promedio de los archivos: 4896.87 KB  
 - Tamaño mínimo de los archivos: 3.58 KB  
 - Tamaño máximo de los archivos: 411418.76 KB  
 - Desviación estándar del tamaño de los archivos: 15877.11 KB  
 - Mediana del tamaño de los archivos: 2440.01 KB  
  
 Conclusiones:  
  
- La mayoría de los archivos PDF analizados contienen texto, lo que facilita su uso directo en modelos de procesamiento de lenguaje natural (NLP) como GPT-4o en el proyecto YoCampo.  
- Los archivos que contienen imágenes requieren procesamiento adicional mediante OCR para convertir el contenido en texto utilizable.  
- Los tamaños de los archivos varían significativamente, desde archivos pequeños de menos de 300 KB hasta archivos grandes que superan los 6000 KB. La desviación estándar relativamente alta indica una gran dispersión en los tamaños de los archivos.