**MODELO TECNICO CONTROLES MINIMOS**



**10/2025**

Índice

[1. Visión General 2](#_Toc212548160)

[2. Características Principales 2](#_Toc212548161)

[3. Instalación y Ejecución 2](#_Toc212548162)

[3.1. Opción A (Recomendada): Usar el Ejecutable (.exe) 2](#_Toc212548163)

[3.2. Opción B: Ejecutar desde el Código Fuente (Para desarrolladores) 2](#_Toc212548164)

[4. Estructura de Carpetas 2](#_Toc212548165)

[4.1. Estructura de Entrada 2](#_Toc212548166)

[4.2. Estructura de Salida 2](#_Toc212548167)

[5. Flujo de Ejecución 2](#_Toc212548168)

[6. Lógica de Análisis de Tendencias 2](#_Toc212548169)

[6.1. Comparación vs. Mes Anterior 2](#_Toc212548170)

[6.2. Comparación vs. Histórico 2](#_Toc212548171)

[6.3. Clasificación de Tendencia 2](#_Toc212548172)

[6.4. Caso Especial (Umbral Mínimo de Filas): 2](#_Toc212548173)

[7. Archivos de Salida Detallados 2](#_Toc212548174)

[7.1. resumen\_MMYYYY.xlsx 2](#_Toc212548175)

[7.2. historico\_completo.xlsx 2](#_Toc212548176)

[8. Diccionario 2](#_Toc212548177)

[8.1. Componentes Principales 2](#_Toc212548178)

[8.2. Estructura y Archivos (Entrada) 2](#_Toc212548179)

[8.3. Estructura y Archivos (Salida) 2](#_Toc212548180)

[8.4. Lógica de Análisis y Clasificación 2](#_Toc212548181)

[8.5. Instalación y Ejecución 2](#_Toc212548182)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Controles mínimos aplicados a GSTI** | | | | |
| Solicitante: | | | | Felipe Acuña |
|  | | |  | |
| Versión: | 1.0 | Fecha: 24-10-2025 | | Ejecutor: Dariac Saldias |
| Versión: | 2.0 | Fecha: 28-10-2025 | | Ejecutor: Dariac Saldias |
|  | | | | |
| Detalle: | 1.0 | Estructura principal e información de la documentación | | |
| Detalle: | 2.0 | Ajuste de estructura profesional de documentación | | |

# **1. Visión General**

El **Analizador de Tendencia de Archivos** es una aplicación de escritorio desarrollada en Python con una interfaz gráfica (GUI) creada con Tkinter. Su objetivo principal es automatizar el proceso de análisis de grandes volúmenes de archivos distribuidos en carpetas mensuales, identificando patrones y tendencias en su tamaño y cantidad de registros a lo largo del tiempo.

La herramienta está diseñada para procesar una estructura de carpetas específica, donde cada subcarpeta representa un mes en formato MMYYYY. El sistema extrae, lee y analiza cada archivo, generando reportes detallados por mes y un informe histórico consolidado que permite una visión macro de la evolución de los datos.

# **2. Características Principales**

* **Procesamiento Automatizado:** Analiza una carpeta base que contiene subcarpetas mensuales (MMYYYY) con un solo clic.

Interfaz de usuario gráfica, Aplicación, Tabla

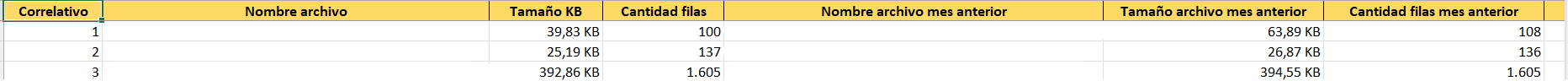
El contenido generado por IA puede ser incorrecto.

* **Interfaz Gráfica Intuitiva:** Facilita la selección de la carpeta de trabajo, el inicio y la cancelación del proceso, y la visualización del progreso en tiempo real.

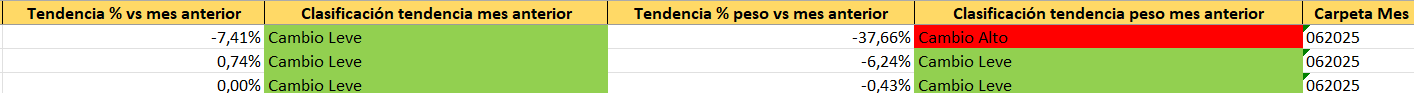
Interfaz de usuario gráfica, Texto, Aplicación

El contenido generado por IA puede ser incorrecto.

* **Manejo de Archivos Comprimidos:** Descomprime automáticamente archivos en formatos .zip, .tar, .gz, .bz2, .7z y analiza su contenido de forma transparente. Admite la descompresión anidada.
* **Soporte Multi-Formato:** Lee y procesa una amplia variedad de formatos de archivo, incluyendo .xlsx, .xls, .csv, .json, .parquet, .txt y .html.
* **Análisis de Tendencias Avanzado:**
  + Compara cada archivo con su versión del **mes anterior**, calculando la variación porcentual en filas y tamaño.
  + Compara cada archivo con su **promedio histórico**, ofreciendo una perspectiva a largo plazo.
  + Utiliza un algoritmo de similitud de nombres (SequenceMatcher) para encontrar archivos correspondientes incluso si sus nombres varían ligeramente.



* **Clasificación de Cambios:** Etiqueta automáticamente las variaciones como "Cambio Leve", "Cambio Moderado", "Cambio Alto" o "Cambio Crítico" según umbrales predefinidos, facilitando la identificación de anomalías.



* **Generación de Reportes:**
  + Crea una carpeta resumen\_generado con **resumen mensual en Excel (.xlsx)** para cada carpeta procesada, con formato condicional para una fácil lectura.

Interfaz de usuario gráfica, Texto

El contenido generado por IA puede ser incorrecto.

* + Crea un reporte de **detalle de nulos (detalle\_nulos.csv)** que identifica las filas y columnas con datos vacíos o nulos en cada archivo.



* + Genera y mantiene un **archivo histórico consolidado** en la carpeta raíz **(historico\_completo.xlsx)** con todos los datos procesados.

Tabla

El contenido generado por IA puede ser incorrecto.

* **Manejo Inteligente de Recursos:**
  + Implementa modos de lectura eficientes para archivos grandes para evitar el consumo excesivo de memoria RAM.
  + Maneja errores de "espacio insuficiente en disco", permitiendo al usuario liberar espacio y continuar el proceso.
* **Gestión de Duplicados:** Antes de iniciar, detecta reportes generados previamente y ofrece moverlos a una carpeta de respaldo para evitar conflictos y duplicidad de datos.

# **3. Instalación y Ejecución**

Existen dos maneras de utilizar la aplicación, dependiendo de si prefieres la simplicidad de un ejecutable o la flexibilidad de ejecutar el código fuente directamente.

## **3.1. Opción A (Recomendada): Usar el Ejecutable (.exe**)

Se ha generado un archivo ejecutable (.exe) que empaqueta la aplicación y todas sus dependencias. Este es el método más sencillo y no requiere ninguna instalación previa.

* **Ventajas:**
  + No es necesario tener Python instalado en el sistema.
  + No se requiere la instalación manual de bibliotecas.
  + Es portable y fácil de distribuir.
* **Instrucciones:**
  1. Localiza el archivo ejecutable (cm.exe).



Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

El contenido generado por IA puede ser incorrecto.

* 1. Haz doble clic sobre él para iniciar la aplicación directamente.
  2. (Opcional). Para más fácil acceso:
* Click derecho en cm.exe
* Crear acceso directo. (Mostrar mas opciones en Windows 11)
* Mover este acceso directo al escritorio(No borrar ni manipular la carpeta principal).

Interfaz de usuario gráfica, Texto, Aplicación

El contenido generado por IA puede ser incorrecto.

## **3.2. Opción B: Ejecutar desde el Código Fuente (Para desarrolladores)**

Si prefieres o necesitas ejecutar el script directamente, debes tener un entorno de Python configurado.

* **Requisitos:**
  + Python 3.x instalado.
  + Las siguientes bibliotecas de Python:
    - pandas: Para la manipulación y análisis de datos.
    - openpyxl: Para leer y escribir archivos Excel (.xlsx).
    - xlrd: Para leer formatos antiguos de Excel (.xls).
    - py7zr: (Opcional) Necesario únicamente si necesitas descomprimir archivos con formato .7z.
* Instalación de Dependencias: Puedes instalar todas las bibliotecas necesarias ejecutando el siguiente comando en tu terminal:

pip install -r requirements.txt

* Ejecución: Una vez instaladas las dependencias, ejecuta el script.

# **4. Estructura de Carpetas**

## **4.1. Estructura de Entrada**

La herramienta requiere que los archivos estén organizados en una carpeta base. Dentro de esta, debe existir una subcarpeta por cada mes a analizar, utilizando el formato MMYYYY.

/Mi\_Carpeta\_Base/

├── 012025/

│ ├── archivo1.csv

│ ├── datos\_enero.zip

│ └── subcarpeta/

│ └── otro\_archivo.xlsx

├── 022025/

│ ├── archivo1\_v2.csv

│ ├── datos\_febrero.zip

│ └── ...

└── 032025/

├── archivo1.csv

└── ...

## **4.2. Estructura de Salida**

La aplicación genera los siguientes archivos y carpetas:

1. **Dentro de cada carpeta mensual (MMYYYY):**
   * Se crea una nueva carpeta llamada resumen\_generado.
   * Dentro de resumen\_generado se guardan:
     + resumen\_MMYYYY.xlsx: El reporte detallado para ese mes.
     + detalle\_nulos.csv: Un listado de todas las celdas vacías encontradas en los archivos de ese mes.
2. **Dentro de la carpeta base (Mi\_Carpeta\_Base):**
   * historico\_completo.xlsx: Un archivo Excel que consolida la información clave de todos los meses procesados.
   * historico\_completo.csv: La misma información en formato CSV.
   * process.log: Un archivo de registro técnico para depuración.

# **5. Flujo de Ejecución**

1. **Inicio de la Aplicación:** El usuario ejecuta el script y se abre la interfaz gráfica.
2. **Selección de Carpeta:** El usuario selecciona la carpeta base que contiene las subcarpetas mensuales. La ruta seleccionada se guarda para futuros usos.
3. **Verificación de Duplicados:** Antes de procesar, el sistema busca archivos de reporte (historico\_completo.xlsx) o carpetas (resumen\_generado) existentes. Si los encuentra, pregunta al usuario si desea moverlos a una carpeta de respaldo en el Escritorio.
4. **Inicio del Proceso:** El usuario hace clic en "Iniciar procesamiento".
5. **Listado de Meses:** El script identifica todas las subcarpetas con formato MMYYYY y las ordena cronológicamente.
6. **Procesamiento Iterativo por Mes:**
   * Para cada carpeta de mes, el script realiza las siguientes acciones:
     + **Expansión de Archivos:** Busca todos los archivos, incluyendo el contenido de los archivos comprimidos. Los archivos se extraen a una carpeta temporal en el Escritorio.
     + **Lectura y Análisis:** Lee cada archivo para contar sus filas válidas (no vacías) y obtener su tamaño. Si un archivo no puede ser leído como una tabla, solo se registra su tamaño.
     + **Análisis de Nulos:** Identifica y registra las filas y columnas con valores nulos o vacíos.
     + **Cálculo de Tendencias:** Compara cada archivo con los datos del mes anterior y con el histórico (si existen) para calcular las variaciones.
     + **Generación de Reporte Mensual:** Guarda los resultados en el archivo resumen\_MMYYYY.xlsx dentro de la carpeta resumen\_generado.
7. **Actualización del Histórico:** Al finalizar cada mes, la información relevante se agrega al DataFrame histórico en memoria.
8. **Guardado del Histórico Final:** Una vez procesados todos los meses, el script guarda el DataFrame histórico completo en los archivos historico\_completo.xlsx y historico\_completo.csv en la carpeta base.
9. **Limpieza:** Se eliminan todas las carpetas temporales creadas durante el proceso.
10. **Finalización:** La interfaz muestra un mensaje de "Proceso finalizado" y las barras de progreso se completan.

# **6. Lógica de Análisis de Tendencias**

El núcleo del análisis reside en la función worker\_procesar. La lógica de comparación es la siguiente:

## **6.1. Comparación vs. Mes Anterior**

* Para cada archivo del mes actual (ej. 022025), el sistema busca un archivo correspondiente en el mes anterior (ej. 012025).
* La búsqueda se realiza en dos pasos:
  1. **Coincidencia Exacta:** Busca un archivo con exactamente el mismo nombre.
  2. **Coincidencia por Similitud:** Si no hay coincidencia exacta, calcula un **ratio de similitud** entre el nombre del archivo actual y todos los del mes anterior. Si encuentra uno con una similitud superior al 80% (>= 0.8), lo considera su contraparte.
* Si se encuentra una contraparte, se calcula el cambio porcentual:
  1. Tendencia % = ((Valor\_Actual - Valor\_Anterior) / Valor\_Anterior) \* 100
* Este cálculo se realiza tanto para la **cantidad de filas** como para el **tamaño en KB**.

## **6.2. Comparación vs. Histórico**

* De manera similar, para cada archivo, se buscan todas sus apariciones en los meses previos (excluyendo el mes actual).
* Se calcula el **promedio** de la cantidad de filas y del tamaño de todas esas apariciones históricas.
* Se calcula el cambio porcentual del valor actual respecto a ese promedio histórico.

## **6.3. Clasificación de Tendencia**

La función clasificar\_tendencia asigna una categoría al cambio porcentual calculado:

* **Cambio Leve:** Variación de hasta ±10%.
* **Cambio Moderado:** Variación entre ±10% y ±30%.
* **Cambio Alto:** Variación entre ±30% y ±50%.
* **Cambio Crítico:** Variación superior a ±50%.

## **6.4. Caso Especial (Umbral Mínimo de Filas):**

Para evitar que pequeñas variaciones en archivos muy chicos (ej. de 2 a 4 filas) sean clasificadas como "Cambio Crítico" (+100%), existe un **umbral configurable**. Si el archivo actual o su referencia tienen menos filas que este umbral (por defecto 10), la tendencia se clasifica automáticamente como **"Cambio Leve"**, sin importar el porcentaje.

# **7. Archivos de Salida Detallados**

## **7.1. resumen\_MMYYYY.xlsx**

| **Columna** | **Descripción** |
| --- | --- |
| **Correlativo** | Identificador numérico único dentro del mes. |
| **Nombre archivo** | Ruta relativa del archivo dentro de la carpeta del mes. |
| **Tamaño KB** | Tamaño del archivo en Kilobytes. |
| **Cantidad filas** | Número de filas con datos (excluyendo filas completamente vacías). |
| **Nombre archivo mes anterior** | Nombre del archivo coincidente en el mes previo. |
| **Tamaño archivo mes anterior** | Tamaño en KB del archivo del mes previo. |
| **Cantidad filas mes anterior** | Cantidad de filas del archivo del mes previo. |
| **Tendencia % vs mes anterior** | Variación porcentual de filas respecto al mes anterior. |
| **Clasificación tendencia mes anterior** | Categoría del cambio (Leve, Moderado, Alto, Crítico). |
| **Tendencia % peso vs mes anterior** | Variación porcentual del tamaño respecto al mes anterior. |
| **Clasificación tendencia peso mes anterior** | Categoría del cambio de tamaño. |
| **Carpeta Mes** | Carpeta del mes a la que pertenece el registro (MMYYYY). |

## **7.2. historico\_completo.xlsx**

| **Columna** | **Descripción** |
| --- | --- |
| **Correlativo, Nombre archivo, Tamaño KB, Cantidad filas, Carpeta Mes** | Igual que en el resumen mensual. |
| **Tendencia histórica %** | Variación porcentual de filas respecto al promedio histórico. |
| **Clasificación tendencia histórica** | Categoría del cambio histórico. |
| **Tendencia histórica % peso** | Variación porcentual del tamaño respecto al promedio histórico. |
| **Clasificación tendencia peso histórica** | Categoría del cambio de tamaño histórico. |

# **8. Diccionario**

## **8.1. Componentes Principales**

* **Analizador de Tendencia de Archivos**: Aplicación de escritorio (Python/Tkinter) diseñada para automatizar el análisis de archivos en carpetas mensuales (MMYYYY), identificando tendencias en tamaño y cantidad de registros.
* **Gestión de Duplicados**: Característica que detecta reportes generados previamente al inicio y ofrece moverlos a una carpeta de respaldo para evitar conflictos.
* **Manejo de Archivos Comprimidos**: Capacidad de la herramienta para descomprimir automáticamente formatos .zip, .tar, .gz, .bz2, y .7z, incluyendo descompresión anidada.
* **Soporte Multi-Formato**: Capacidad de leer y procesar archivos .xlsx, .xls, .csv, .json, .parquet, .txt, y .html.
* **worker\_procesar**: Función principal que contiene la lógica central del análisis de tendencias.

## **8.2. Estructura y Archivos (Entrada)**

* **Carpeta Base**: Directorio principal seleccionado por el usuario que contiene las subcarpetas mensuales a procesar.
* **MMYYYY**: Formato requerido para las subcarpetas que representan cada mes (ej. "012025").

## **8.3. Estructura y Archivos (Salida)**

* **resumen\_generado**: Nombre de la subcarpeta creada dentro de cada carpeta mensual (MMYYYY) para almacenar los reportes de ese mes.
* **resumen\_MMYYYY.xlsx**: Reporte en Excel generado para un mes específico, que contiene el análisis detallado de los archivos de ese período.
* **detalle\_nulos.csv**: Reporte generado por mes que lista las filas y columnas específicas donde se encontraron datos vacíos o nulos.
* **historico\_completo.xlsx**: Archivo Excel consolidado (ubicado en la carpeta base) que acumula la información clave de todos los meses procesados.
* **historico\_completo.csv**: Versión en formato CSV del archivo histórico consolidado.
* **process.log**: Archivo de registro técnico (log) utilizado para depuración.

## **8.4. Lógica de Análisis y Clasificación**

* **Análisis de Tendencias**: Proceso de comparar cada archivo con su versión del mes anterior y con su promedio histórico para calcular variaciones porcentuales.
* **SequenceMatcher**: Algoritmo de la biblioteca difflib de Python, utilizado para encontrar archivos correspondientes entre meses basándose en la similitud de nombres (>= 80% o 0.8).
* **Comparación vs. Mes Anterior**: Cálculo de la variación porcentual en filas y tamaño de un archivo actual contra su contraparte del mes previo (identificada por coincidencia exacta o por similitud).
* **Comparación vs. Histórico**: Cálculo de la variación porcentual en filas y tamaño de un archivo actual contra el *promedio* de todas sus apariciones en meses anteriores.
* **Umbral Mínimo de Filas**: Lógica de excepción (por defecto 10 filas) que clasifica una variación como "Cambio Leve", independientemente del porcentaje, si el archivo actual o su referencia son muy pequeños.
* **Clasificación de Cambios**: Etiquetas asignadas a las variaciones porcentuales:
  + **Cambio Leve**: Variación de hasta ±10%.
  + **Cambio Moderado**: Variación entre ±10% y ±30%.
  + **Cambio Alto**: Variación entre ±30% y ±50%.
  + **Cambio Crítico**: Variación superior a ±50%.

## **8.5. Instalación y Ejecución**

* **cm.exe**: El archivo ejecutable (Opción A) que empaqueta la aplicación y sus dependencias, permitiendo su uso sin instalar Python.
* **requirements.txt**: Archivo de texto que lista las dependencias de Python (pandas, openpyxl, xlrd, py7zr) necesarias para ejecutar la aplicación desde el código fuente (Opción B).