

Caso de éxito

Automatización del análisis y distribución de partidas conciliatorias mediante procesamiento de datos en Python

Programa de Fortalecimiento de Habilidades y Herramientas de Inteligencia Artificial para el Sector Público

Nombre Autor (a): Diana Johanna Bastidas Devia
2025

1. Nombre del Caso/Iniciativa

Automatización del análisis y distribución de partidas conciliatorias mediante procesamiento de datos en Python

2. Entidad(es) Responsable(s)

Contaduría General de la Nación, Git Empresas

3. Sector Administrativo (Salud, Justicia, Educación, etc.)

Hacienda y finanzas públicas del Estado

4. Área de Aplicación

Fortalecimiento de la Gestión Interna y Procesos de Apoyo

5. Problema Público Abordado

La tarea a optimizar consistía en el tratamiento del reporte maestro de conciliación generado por el sistema CHIP, un archivo de Excel voluminoso, no segmentado y con todas las inconsistencias del país. Cada analista debía filtrar manualmente el archivo para identificar las entidades y transacciones que le correspondían, localizar las contrapartes y construir su propia hoja de trabajo. Este procedimiento exigía búsquedas visuales reiteradas y cruces manuales, resultaba lento, generaba alta carga operativa y se asociaba con riesgos de errores por omisión o por cruce equivocado. Ese tiempo invertido en preparar datos restaba capacidad para el análisis sustantivo de las partidas conciliatorias, lo que afectaba la oportunidad y la calidad del proceso de conciliación contable.

6. Solución de IA Implementada

La solución se diseñó como un esquema híbrido que combina Inteligencia Artificial generativa para apoyar el desarrollo y automatización de procesos para ejecutar la tarea. La IA (Gemini) se utilizó como asistente de programación para co-crear, depurar y optimizar una serie de scripts en Python; estos scripts realizan el procesamiento masivo del reporte maestro de CHIP.

En el flujo operativo, primero el script se autentica con Google Drive y carga el archivo maestro Base_Conciliacion.xlsx en un DataFrame de Pandas, omitiendo las filas de

encabezado que no contienen datos útiles. Después limpia y transforma los datos: crea un índice único por fila, normaliza las columnas críticas mediante funciones de cadena y genera campos tokenizados que reducen inconsistencias de formato. Luego separa las filas que actúan como “padre” (con el saldo global) de las filas “hijo” asociadas a las personas analistas, y construye la lógica de cruce mediante un auto-merge de la base de entidades utilizando el Id de transacción como llave.

A partir de ese cruce, el script arma una base maestra de “líder”, aplica el filtro de calibración por índice (index_1 distinto de index_2) para obtener todas las interacciones válidas entre entidades y clasifica las interacciones según reglas de negocio, como la presencia de “gobierno” en el marco normativo para marcar o no la validez en una columna Filtro_Valido. Finalmente, la herramienta genera dos tipos de productos: un reporte líder general con todas las interacciones y su clasificación, y una serie de reportes individuales, uno por cada analista, con las filas válidas ya filtradas, listos para el trabajo de análisis.

7. Tecnologías Utilizadas (ej. PLN, Visión por Computador, Machine Learning)

La solución se apoya en Inteligencia Artificial generativa y en herramientas de procesamiento de datos. Como herramienta de IA se utilizó Gemini, que actuó como asistente de desarrollo para escribir, corregir y optimizar el código en Python, incluida la lógica de tokenización, los algoritmos de cruce y la estructuración del flujo ETL. El procesamiento de datos se realizó con Python y la biblioteca Pandas para la ingesta desde Excel, la limpieza, las transformaciones y los cruces; Google Colab sirvió como entorno de ejecución en la nube, y Google Drive como repositorio centralizado para el archivo maestro Base_Conciliacion.xlsx y para los reportes generados.

8. Resultados

Antes de la solución, la preparación y distribución de la información tomaba entre dos y tres horas por analista, porque cada persona debía extraer manualmente sus Id de transacción, identificar las contrapartes y armar su propia base de trabajo. Después de la automatización, el liderazgo del equipo ejecuta un único script de Python en Google Colab y, en menos de tres minutos, obtiene el reporte líder general y todos los reportes individuales prefiltrados y listos para el análisis. Esto supone un ahorro muy

significativo de tiempo: la tarea que antes se realizaba analista por analista pasa a completarse de forma centralizada, con una reducción del orden de horas de trabajo a pocos minutos.

En términos de calidad, el diseño de la lógica de cruce y el uso de índices únicos permiten eliminar los errores de omisión y de emparejamiento manual: el documento señala que la lógica se calibró frente a un caso de prueba conocido (las 314 interacciones correspondientes al analista DBASTIDAS), logrando que el reporte automatizado reproduzca de manera completa ese universo, lo que equivale a una reducción del 100% de los errores de omisión derivados del procedimiento manual. Además, las personas analistas ya no reciben datos crudos, sino reportes accionables, con las interacciones gobierno–gobierno identificadas y la contraparte visible, lo que mejora la calidad del trabajo y permite concentrar el esfuerzo en el análisis técnico de las diferencias.

9. Factor de Sostenibilidad y Escalabilidad

La solución incide directamente en la misión de la Contaduría General de la Nación, porque acelera la identificación y resolución de partidas conciliatorias y mejora la fiabilidad de los saldos por conciliar en los reportes contables del sector público. Al liberar al personal analista de tareas repetitivas y de baja complejidad, la entidad puede dedicar más tiempo al análisis crítico y a la mitigación de riesgos operativos asociados a la contabilidad pública, lo que fortalece la calidad y la oportunidad de la información que se pone al servicio de la gestión estatal y de la ciudadanía. En cuanto a sostenibilidad y escalabilidad, el caso subraya que la solución se construye exclusivamente con herramientas de costo cero: Python, Pandas, Matplotlib, Google Colab, Google Drive y la IA generativa de Gemini. Esto elimina barreras presupuestarias y permite que cualquier área con una lógica de conciliación similar replique el enfoque. La persona nueva solo necesita acceso al notebook de Colab y a la carpeta de trabajo; el script puede procesar desde unos pocos miles hasta cientos de miles de filas con el mismo flujo operativo, dado que la infraestructura en la nube asume la carga. Además, el reporte líder generado funciona como base estructurada para desarrollar, en fases posteriores, tableros de inteligencia de negocios y otros instrumentos de monitoreo que potencian aún más la gestión.