

Caso de éxito

**Aceleración en la Generación de
Insights Fiscales mediante
Inteligencia Artificial.**

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

HUGO ALCIDES PEREZ PINILLA
2025

1. Nombre del Caso/Iniciativa

Aceleración en la Generación de Insights Fiscales mediante Inteligencia Artificial

2. Entidad(es) Responsable(s)

DIAN- OFICINA DE SEGURIDAD DE LA INFORMACIÓN.

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

Gobierno - Tributario

4. Área de Aplicación

Soporte a la Toma de Decisiones y Planificación Estratégica

5. Problema Público Abordado

El proceso administrativo para optimizar es el análisis de los reportes de recaudación de impuestos, una tarea fundamental para la toma de decisiones en la entidad. Periódicamente, los analistas reciben grandes volúmenes de datos en formatos de hoja de cálculo (CSV o Excel) que debían ser procesados para generar informes y visualizaciones.

El problema principal era la excesiva carga de trabajo manual y la lentitud del proceso. Los analistas deben invertir horas, o incluso días, en tareas repetitivas: limpiar los datos, calcular estadísticas descriptivas, crear tablas dinámicas y diseñar gráficos manualmente en Excel. Este método no solo era ineficiente, sino también propenso a errores humanos.

Esta lentitud creaba un cuello de botella significativo. Para cuando los informes llegaban a los directivos, la información podía haber perdido relevancia. Además, la capacidad de exploración era muy limitada; si un directivo tenía una pregunta nueva sobre los datos, debía solicitar un nuevo análisis que volvía a iniciar el ciclo manual. Esto dificultaba la identificación ágil de tendencias, anomalías o la comparación efectiva del rendimiento entre diferentes seccionales o periodos fiscales, limitando la capacidad de la organización para reaccionar y planificar estratégicamente.

6. Solución de IA Implementada

Para resolver el reto de la lentitud y el trabajo manual, se desarrolló una aplicación web que automatiza el análisis de datos de recaudación utilizando la API de Google Gemini. La solución se implementó en un proceso de cuatro pasos automáticos que se activan en cuanto el usuario sube un archivo.

El proceso que se siguió fue el siguiente:

1. **Análisis Descriptivo Automatizado:** Tras cargar un archivo (CSV o Excel), la aplicación envía una muestra de los datos al modelo Gemini 2.5 Pro. Se utilizó un prompt específico que instruye a la IA para que actúe como un analista de datos experto. El prompt solicita un resumen ejecutivo de los datos (periodos, valores clave, etc.) y un detalle técnico de cada columna. Gracias a la capacidad de Gemini para generar respuestas en formato JSON estructurado, la aplicación recibe esta información y la presenta al usuario de forma inmediata.
2. **Análisis de Series Temporales:** Si la IA detecta una columna de fechas en el paso anterior, automáticamente realiza una segunda llamada al modelo Gemini 2.5 Flash. Se le pide que analice tendencias, estacionalidad y genere un breve pronóstico. Este análisis, que antes tomaba horas, ahora se presenta en segundos.
3. **Sugerencia Inteligente de Gráficos:** Basándose en el resumen inicial, la aplicación realiza una tercera llamada a Gemini. El prompt solicita de 3 a 5 sugerencias de visualizaciones relevantes para datos fiscales, como "Recaudación por seccional" o "Evolución anual de impuestos". La IA devuelve la configuración exacta para cada gráfico (tipo, columnas a usar, título), que se presentan como botones interactivos.
4. **Generación de Gráficos por Lenguaje Natural:** Finalmente, se le dio al usuario el control directo. A través de una caja de texto, el analista puede escribir peticiones como "muéstrame un gráfico de pastel del recaudo por tipo de impuesto". Esta solicitud, junto con el contexto de los datos, se envía a Gemini 2.5 Pro, que la traduce en la configuración JSON necesaria para renderizar el gráfico solicitado al instante.

Este enfoque busca transformar un flujo de trabajo de días en una experiencia interactiva de minutos, eliminando tareas repetitivas y permitiendo a los analistas centrarse en la interpretación estratégica de la información.

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

Los recursos utilizados fueron:

1. API de Google Gemini: Es el único componente con un costo potencial. Sin embargo, Google ofrece un generoso nivel de uso gratuito, más que suficiente para el desarrollo, la experimentación y el uso a pequeña escala. Más allá de eso, opera en un modelo de pago por uso, lo que lo mantiene como una opción de bajo costo en comparación con otras soluciones de IA que requieren entrenamiento o infraestructura propia.
2. Librerías de Desarrollo (Frontend): Se utilizaron React, TypeScript y Recharts para la interfaz y los gráficos. Todas estas herramientas son de código abierto y completamente gratuitas
3. Herramientas de Estilo y de Datos: Se usó Tailwind CSS para el diseño y la librería `xlsx` para leer archivos de Excel. Ambas se incluyeron a través de CDNs públicos y gratuitos, eliminando la necesidad de procesos de compilación complejos
4. Entorno de Desarrollo y Despliegue: El desarrollo solo requirió un editor de código estándar (como VS Code) y un navegador web. Al ser una aplicación de archivos estáticos, puede ser alojada de forma gratuita o a un costo muy bajo en numerosas plataformas (como GitHub Pages o Vercel).

8. Resultados.

La implementación del "Analizador IA de Recaudo de Impuestos" ha supuesto una transformación radical en el flujo de trabajo del análisis de datos, pasando de un proceso laborioso y lento a una experiencia ágil e interactiva.

El cambio más significativo se observa en el **ahorro de tiempo**. Antes, el proceso completo —desde la recepción de los datos brutos hasta la entrega de un informe con estadísticas y gráficos— podía tomar entre 8 y 16 horas de trabajo manual por parte de un analista. Con la nueva aplicación, el mismo ciclo se completa en **menos de 5 minutos**. Esto incluye la carga del archivo, el análisis descriptivo y de series temporales por parte de la IA, y la generación de un panel interactivo con múltiples visualizaciones listas para explorar. Esto representa una **optimización del tiempo de trabajo superior al 95%**, liberando a los analistas para que se concentren en la interpretación estratégica en lugar de en la preparación de datos.

En cuanto a la **reducción de errores**, la solución ha sido igualmente impactante. El proceso manual anterior era susceptible a errores de transcripción, fórmulas incorrectas en las hojas de cálculo o inconsistencias en la agregación de datos. Al automatizar el análisis directamente desde la fuente de datos, **se han eliminado prácticamente todos los errores operativos**. La IA realiza los cálculos y la generación de gráficos de manera consistente y estandarizada, garantizando que las visualizaciones reflejen fielmente la información del archivo original.

Finalmente, la **mejora en la calidad y la capacidad analítica** es profunda. Los informes generados no solo son más rápidos y precisos, sino también mucho más completos desde el inicio, incluyendo análisis de tendencias que antes requerían un esfuerzo adicional considerable. La verdadera revolución, sin embargo, reside en la capacidad de exploración dinámica. Antes, cualquier pregunta nueva o la necesidad de un gráfico diferente iniciaba otro ciclo de trabajo manual. Ahora, gracias a la función de "Gráfico Personalizado", cualquier usuario puede formular preguntas en lenguaje natural y obtener visualizaciones al instante. Esto ha democratizado el acceso a la información, permitiendo a los directivos y analistas pasar de un modelo de reportes estáticos a una plataforma de descubrimiento de datos en tiempo real, fomentando una toma de decisiones más ágil e informada.

9. Factor de Sostenibilidad y Escalabilidad

La replicación y escalado de esta solución son notablemente sencillos, diseñados precisamente para ser adoptados con una barrera de entrada muy baja.

Facilidad para Replicar: La solución es extremadamente fácil de implementar para un colega en otra área o entidad. Al no requerir una infraestructura de servidor compleja, bases de datos o la configuración de modelos de machine learning propios, el despliegue es inmediato. La aplicación es un cliente web autónomo que delega todo el procesamiento de inteligencia artificial a la API de Gemini.

Un colega podría tener una copia funcional de la aplicación en cuestión de minutos. El proceso es tan simple como servir un conjunto de archivos estáticos (HTML, CSS, JavaScript) desde cualquier servidor web básico.

¿Qué se Necesitaría para Empezar?

1. El Código Fuente: Acceso a los archivos de la aplicación (proporcionados en este caso).
2. Una Clave de API de Google Gemini: Este es el único requisito fundamental. La persona interesada solo necesita generar su propia clave gratuita desde Google AI Studio e insertarla en la configuración.
3. Datos Propios: Tener a la mano los archivos CSV o Excel que deseen analizar.

No se necesita experiencia en machine learning ni en administración de sistemas. Con conocimientos básicos de desarrollo web, cualquier persona puede no solo implementar la solución, sino también adaptarla, por ejemplo, para analizar diferentes tipos de datos o personalizar la interfaz.

Escalabilidad: La arquitectura es inherentemente escalable. Como el procesamiento intensivo lo realiza la infraestructura de Google, la aplicación puede manejar archivos de datos más grandes o un mayor número de usuarios sin degradar su rendimiento. La escalabilidad no depende de la capacidad de los servidores locales, sino de los límites de uso de la API de Gemini, que son muy generosos y pueden ampliarse fácilmente para casos de uso empresariales.