

Informe Técnico: Agente de IA Analy

28/11/2025



1.0 Introducción

Analy es un agente de Inteligencia Artificial especializado en el análisis y la visualización de datos. Su propósito principal es capacitar a profesionales para que puedan realizar análisis exploratorios complejos sin necesidad de programar. Este informe tiene como objetivo detallar la arquitectura, capacidades funcionales y especificaciones técnicas de Analy, proporcionando una evaluación objetiva de su alcance y aplicabilidad.

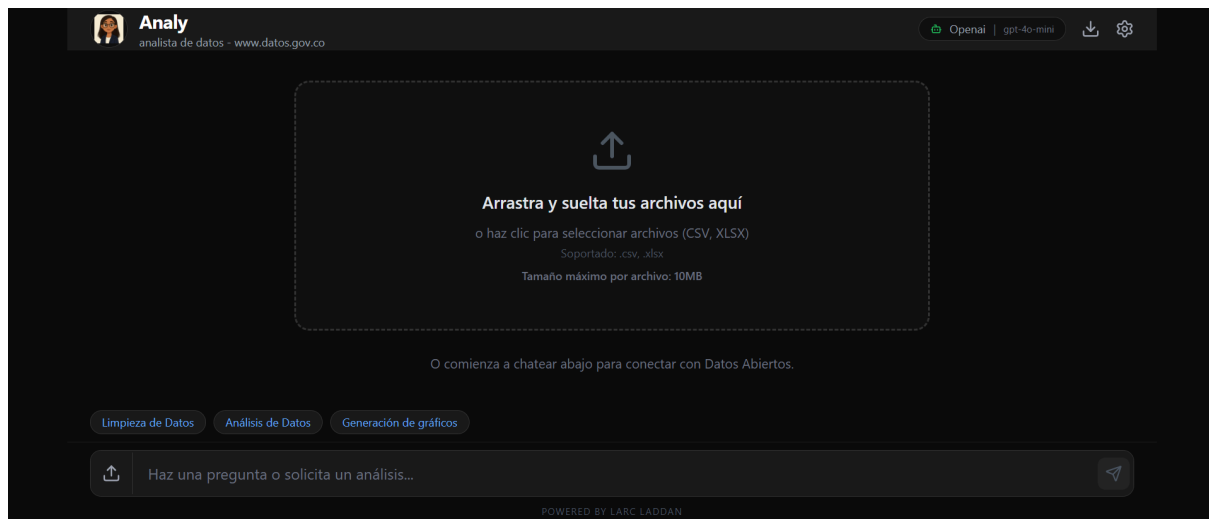
La propuesta de valor central de la herramienta reside en su capacidad para transformar conjuntos de datos brutos en insights accionables y gráficos interactivos, operando a través de la conexión con fuentes de datos públicas y la integración con modelos de IA avanzados. En un entorno donde la toma de decisiones basada en datos es fundamental, Analy se posiciona como un puente estratégico que reduce la brecha entre los datos y la comprensión.

ADVERTENCIA: actualmente Analy es un MVP, y puede cometer errores.

A continuación, se describirá la arquitectura del sistema que sustenta estas capacidades.

2.0 Arquitectura y Componentes Fundamentales

Comprender la arquitectura de Analy es crucial para evaluar su fiabilidad, escalabilidad y seguridad. El sistema está diseñado como una arquitectura híbrida que integra de manera sinérgica fuentes de datos externas y la potencia computacional de modelos de lenguaje de gran escala (LLM) de terceros. Esta estructura modular permite una operación flexible y eficiente.



2.1 Conectividad con Fuentes de Datos

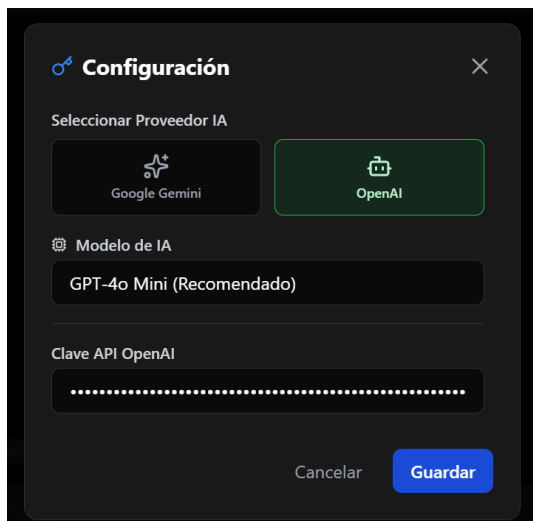
Analy puede acceder a la información a través de dos canales principales, lo que le confiere versatilidad para distintos escenarios de análisis.

- **Fuentes Públicas:** La herramienta establece una conexión directa con el portal de datos abiertos del gobierno de Colombia, www.datos.gov.co. Esta integración le permite acceder no solo a los conjuntos de datos (datasets) alojados en la plataforma, sino también a sus metadatos asociados, facilitando la búsqueda y contextualización de la información.
- **Archivos de Usuario:** Analy ofrece una funcionalidad para que el usuario cargue sus propios archivos de datos. Los formatos de archivo soportados para esta operación son **CSV** y **XLSX**, que son estándares comunes en el manejo de datos tabulares.

2.2 Integración con Modelos de Lenguaje Avanzados

La arquitectura de Analy destaca por su flexibilidad en el núcleo de procesamiento de IA, permitiendo la conexión con múltiples proveedores de modelos de lenguaje.

- Los proveedores de IA explícitamente identificados en la configuración del sistema son **Google** y **OpenAI**.



- El usuario tiene control sobre esta integración, como se observa en la pantalla de "Configuración". Desde allí, es posible seleccionar el proveedor de IA (Google Gemini o OpenAI) y el modelo específico que se desea utilizar para el análisis (ej. GPT-4o Mini), permitiendo adaptar la herramienta a diferentes necesidades de costo y capacidad.

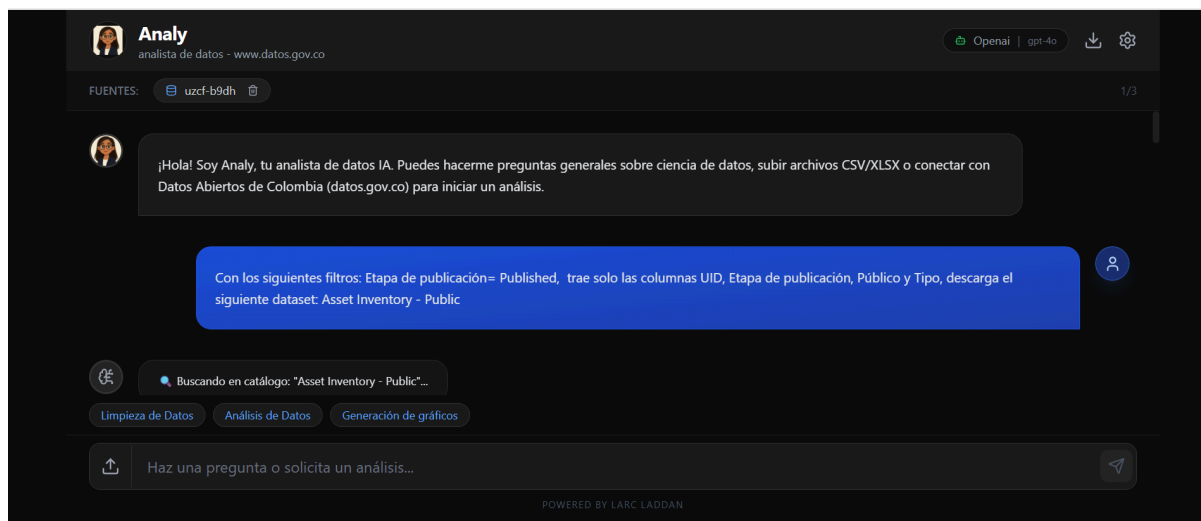
Esta arquitectura de componentes interconectados es la que habilita el robusto conjunto de capacidades funcionales que se detallan en la siguiente sección.

3.0 Capacidades Funcionales

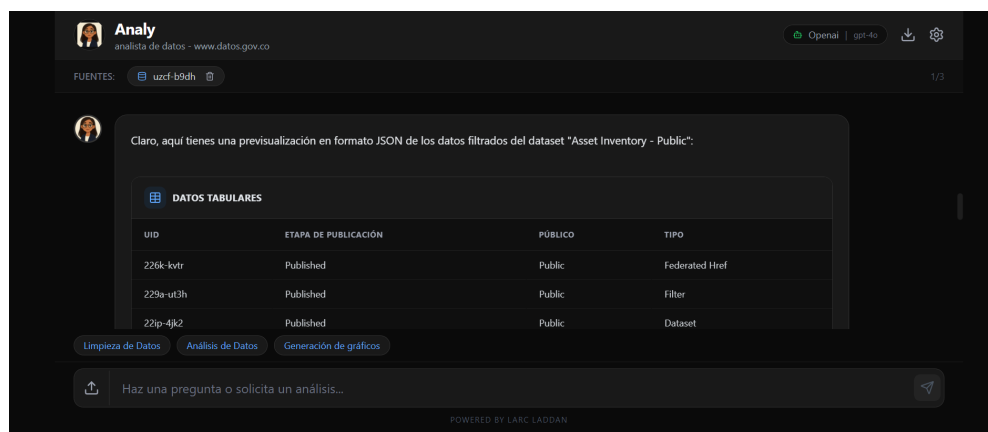
Las capacidades de Analy se enmarcan dentro de un flujo de trabajo de análisis de datos completo, abarcando desde la adquisición y preparación inicial hasta el análisis estadístico y la visualización final de resultados. Todas estas funciones están diseñadas para ser accesibles a través de una interfaz conversacional, donde el usuario dirige el proceso mediante instrucciones en lenguaje natural.

3.1 Adquisición y Preparación de Datos

Analy proporciona herramientas esenciales para la limpieza y preparación de los datos, una etapa crítica en cualquier proyecto de análisis.



1. **Combinación de Fuentes:** La plataforma puede combinar recursos obtenidos directamente desde datos.gov.co con archivos subidos por el usuario, permitiendo enriquecer análisis con datos de distintas procedencias.
2. **Consultas y Filtrado:** Soporta la ejecución de filtros, ordenamientos y consultas mediante el lenguaje **SoSQL**. Esto permite segmentar los datos de manera precisa. Un ejemplo práctico es la solicitud explícita del usuario para filtrar el dataset 'Asset Inventory - Public' por el criterio 'Etapa de publicación=Published', especificando la extracción de las columnas UID, Etapa de publicación, Público y Tipo.
3. **Visualización Preliminar:** Tras una operación de filtrado o consulta, Analy es capaz de mostrar una previsualización de los datos resultantes en un formato tabular claro, como se demuestra en la interfaz con la tabla de "DATOS TABULARES", facilitando la validación inmediata de los resultados.



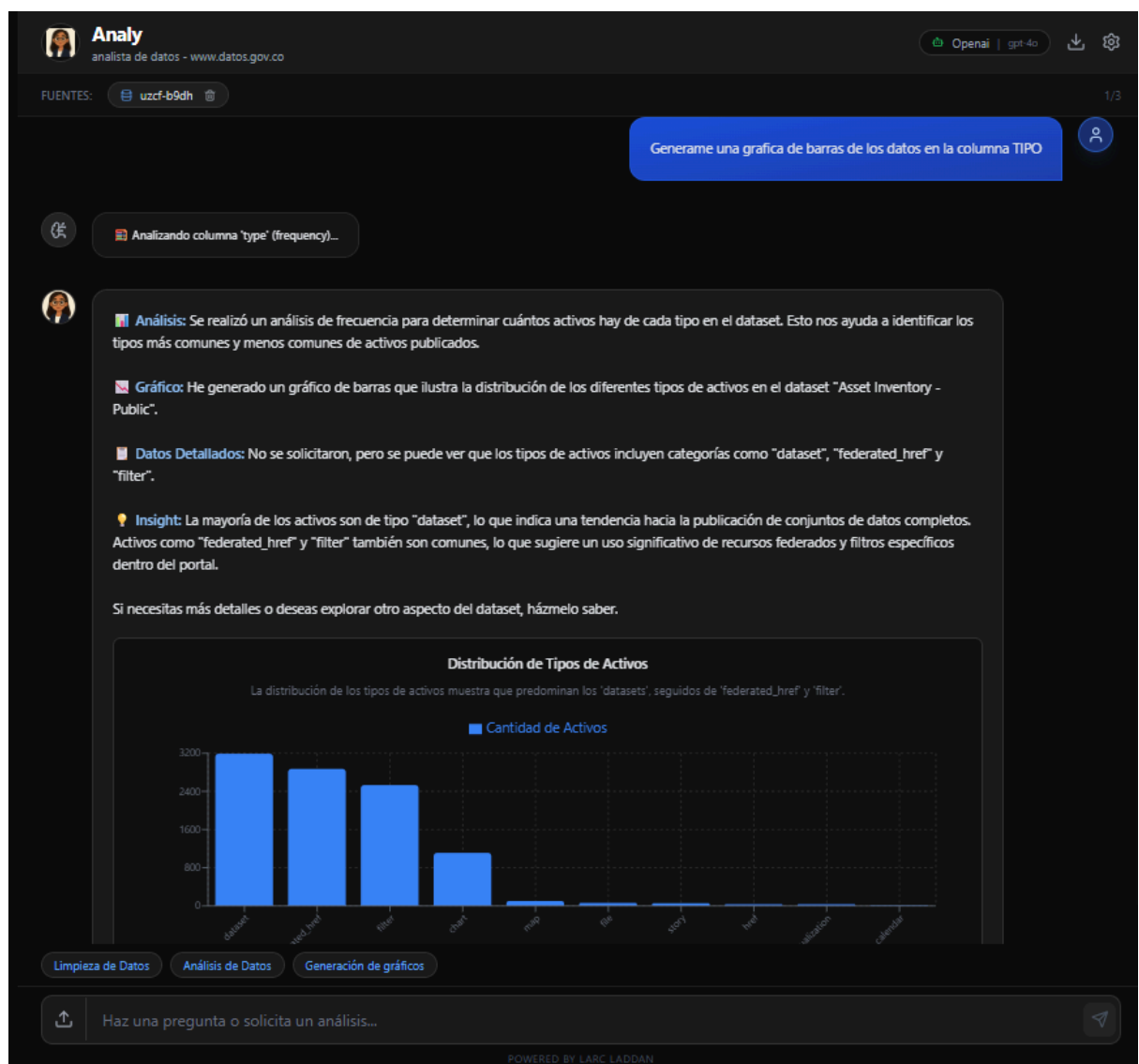
3.2 Análisis Estadístico

Una vez que los datos están preparados, Analy puede ejecutar análisis estadísticos para descubrir patrones y tendencias subyacentes. En el caso de estudio observado, la herramienta realizó un análisis de frecuencia sobre la columna "TIPO" del dataset para determinar la cantidad de activos de cada categoría. Este proceso culminó en la generación de un *insight* clave: "La mayoría de los activos son de tipo 'dataset'", identificando así la tendencia principal en los datos. Adicionalmente, el análisis identificó que los tipos

'federated_href' y 'filter' también son comunes, lo que sugiere un uso significativo de recursos federados y filtros específicos dentro del portal de datos.

3.3 Visualización de Datos

La comunicación de los hallazgos es un componente fundamental del análisis, y Analy lo facilita mediante la generación automática de gráficos. Siguiendo el ejemplo anterior, a partir del análisis de frecuencia, la herramienta generó un gráfico de barras titulado "Distribución de Tipos de Activos". Esta visualización ilustra de forma clara y efectiva la distribución de las categorías de activos, convirtiendo las cifras del análisis en una representación visual comprensible con una leyenda que cuantifica la "Cantidad de Activos" para cada tipo.



3.4 Conciencia Temporal

Una característica notable de Analy es que tiene conocimiento de la fecha y hora actual. Esta capacidad es particularmente útil para realizar análisis de series temporales, aplicar

filtros basados en rangos de fechas recientes o contextualizar datos que tienen una dependencia temporal.

Estas capacidades funcionales demuestran la utilidad de la herramienta, pero para un uso efectivo, es crucial conocer sus límites operativos.

4.0 Especificaciones Técnicas y Limitaciones Operativas

Es importante comprender las limitaciones técnicas de Analy. Estas restricciones no deben ser vistas como debilidades, sino como parámetros de diseño implementados deliberadamente para garantizar la estabilidad del sistema y proteger el rendimiento del navegador del usuario, evitando que la interfaz se congele.

Parámetro	Límite / Especificación
Datasets descargados de datos.gov.co	Hasta 3
Archivos subidos por el usuario	Hasta 3
Formatos de archivo soportados	CSV, XLSX
Tamaño máximo por archivo	10 MB
Límite de filas por dataset	10.000 filas

Nota: El límite de 10.000 filas por dataset es una decisión de diseño para asegurar que el volumen de datos sea "suficiente para análisis estadísticos significativos sin congelar la interfaz" del navegador.

Estas especificaciones definen el marco operativo dentro del cual Analy es más efectivo, lo que nos permite identificar sus casos de uso más adecuados.

5.0 Casos de Uso Potenciales

Las características y especificaciones de Analy la posicionan como una herramienta ideal para perfiles profesionales y tareas que requieren agilidad y accesibilidad en el análisis de datos de tamaño moderado.

- **Análisis Exploratorio de Datos Públicos:** Aprovechando su conexión nativa con datos.gov.co y su motor de visualización sin código, un analista de políticas públicas o un periodista de datos puede utilizar Analy para explorar rápidamente datasets del portal del gobierno colombiano. Podrían identificar tendencias, validar supuestos y generar gráficos preliminares para informes o artículos sin necesidad de escribir una sola línea de código en R o Python.
- **Validación Rápida de Hipótesis:** Gracias a su capacidad para combinar fuentes de datos heterogéneas (archivos locales y datos públicos), un analista de negocios podría cargar un archivo CSV con datos internos de ventas y combinarlo con un dataset demográfico público de datos.gov.co. Esto le permitiría validar de forma ágil una hipótesis de mercado sobre la correlación entre ventas y ciertas características poblacionales a nivel municipal.
- **Generación de Informes y Dashboards Sencillos:** Explotando la capacidad de ingesta de formatos estándar como XLSX y la generación automática de gráficos, un gerente de proyectos puede exportar datos de rendimiento de sus sistemas a un archivo XLSX, cargarlo en Analy y solicitar análisis de frecuencia o distribuciones. Las visualizaciones generadas pueden ser incorporadas directamente en presentaciones de estado para comunicar el progreso y los resultados al equipo o a los stakeholders.

Estos casos de uso demuestran el valor de Analy para democratizar el acceso al análisis de datos, empoderando a usuarios no técnicos para que extraigan valor de la información.

6.0 Ideas complementarias para escalar el proyecto

- Agregar registro y autenticación por usuario.
- Además de archivos CSV y XLSX, podría aceptar archivos TXT, XML, PDF, JSON, DOC.
- Ampliar opciones de modelos y proveedores a conectar via API a Analy.
- Aumentar la capacidad de carga de archivos (cantidad, peso, filas).
- Agregar modo Dark y modo Light.
- Conectar a Analy a otras fuentes de datos, como bases de datos abiertas (Datacommons, wikipedia, arxiv, etc.) o privadas.
- Construir a Analy como un Swarm de agentes especializados.
- Permitir que Analy genere reportes estructurados y estéticos en PDF.
- Robustecer el prompt del agente.

7.0 Conclusión

Este informe ha detallado las características técnicas y funcionales del agente de IA Analy. Se ha establecido que Analy es una herramienta especializada y robusta para el análisis y

la visualización de datos, diseñada con un enfoque en la accesibilidad y la eficiencia. Su arquitectura híbrida, que combina el acceso a datos públicos y privados con la potencia de modelos de IA de vanguardia, le permite ofrecer un flujo de trabajo de análisis completo a través de una interfaz conversacional.

Su impacto principal radica en la reducción de la barrera técnica para el análisis de datos. Al abstraer la complejidad de la programación y la configuración de entornos, Analy se consolida como un activo valioso para profesionales que necesitan obtener insights rápidos y fiables a partir de conjuntos de datos de tamaño moderado, operando dentro de especificaciones técnicas claras que garantizan la estabilidad y una experiencia de usuario fluida.