

Manual del Usuario

Panel Inteligente de Exploración de Datos (EDA)

Tabla de contenidos

1. Introducción
 2. Requisitos previos
 3. Inicio rápido
 4. Barra lateral
 5. Pestaña 1: Ingesta y Procesamiento (ETL)
 6. Pestaña 2: Visualización Dinámica (EDA)
 7. Pestaña 3: Insights de IA
 8. Temas y personalización
 9. Preguntas frecuentes
-

Introducción

El **Panel Inteligente** es una aplicación web interactiva diseñada para realizar análisis exploratorio de datos (EDA) de forma dinámica. Soporta múltiples fuentes de datos, detecta automáticamente el tipo de columnas (numéricas, categóricas, booleanas), identifica problemas de calidad y ofrece visualizaciones interactivas junto con insights generados por IA mediante Groq.

Característica	Descripción
Multi-fuente	Carga datos desde archivos CSV, JSON o URLs
ETL automático	Sanitización, detección de nulos, duplicados y valores atípicos
Clasificación inteligente	Identifica columnas numéricas, categóricas y booleanas automáticamente
Visualizaciones dinámicas	Gráficos Plotly interactivos (histogramas, boxplots, heatmaps, series temporales)
IA integrada	Ánalisis y recomendaciones generadas por Groq (Llama-3)

Requisitos previos

- Navegador web moderno (Chrome, Firefox, Edge, Safari)

- **API Key de Groq** (opcional, para insights con IA) — obtenerla en console.groq.com
 - Si ejecutas localmente: **Python 3.9+** y las dependencias del proyecto
-

Inicio rápido

1. Abre la aplicación (enlace de despliegue o ejecuta `streamlit run app.py` localmente).
 2. En la barra lateral, elige cómo cargar datos: **Archivo CSV**, **Archivo JSON** o **URL**.
 3. Carga tu dataset. La aplicación detectará automáticamente la estructura.
 4. Revisa la pestaña **Ingesta y Procesamiento** para ver la clasificación de columnas y la calidad de los datos.
 5. Explora en **Visualización Dinámica** y genera insights en **Insights de IA** (si tienes API Key configurada).
-

Barra lateral

La barra lateral agrupa la configuración principal de la aplicación.

Carga de datos

Opción	Uso
Archivo CSV	Sube un archivo .csv desde tu dispositivo
Archivo JSON	Sube un archivo .json con estructura tabular
URL	Introduce una URL pública que apunte a un archivo CSV o JSON

Tip: El formato CSV debe usar coma como separador. Se soportan números con formato localizado (ej. 986,101 o 8,874,909).

Filtros y tratamiento de datos

Solo aparecen cuando hay datos cargados:

- **Eliminar filas duplicadas** — Marca para excluir filas duplicadas del análisis.
- **Método de imputación numérica** — Cómo llenar valores faltantes en columnas numéricas:
 - *Sin imputación* — No se modifican.

- *Media* — Se usa la media de la columna.
- *Mediana* — Se usa la mediana.
- *Cero* — Se reemplazan por 0.
- **Eliminar filas con valores atípicos** — Excluye filas con outliers (método IQR).

Columna índice / identificador

Selecciona la columna que actúa como identificador (por ejemplo, ID de cliente). Esta columna no se usa en análisis estadísticos ni visualizaciones.

API Key de Groq

Introduce tu API Key de Groq para habilitar: - **Generar Insights con IA** en la pestaña Insights - Resúmenes con IA en histogramas, boxplots, heatmaps, series temporales, etc. - Chat conversacional sobre el dataset

La API Key se almacena solo en la sesión actual y no se guarda en servidor.

Pestaña 1: Ingesta y Procesamiento (ETL)

Esta pestaña muestra el flujo de limpieza y preparación de datos.

1. Reemplazo de valores vacíos

Las cadenas vacías y espacios en blanco se convierten automáticamente en NaN para un tratamiento consistente.

2. Sanitización y conversión numérica

- **Strings numéricos:** Valores como "986,101" o "8,874,909" se detectan y convierten a numérico.
- **Columnas categóricas:** Se aplican trim (eliminación de espacios) y conversión a minúsculas para unificar categorías.

3. Información del dataset

Se muestra un resumen del número de columnas por tipo:

Tipo	Descripción
Numéricas	Columnas con valores numéricos continuos o enteros
Categóricas	Texto con múltiples categorías
Booleanas	Columnas con solo dos valores (sí/no, 0/1, etc.)

4. Detección de calidad

- **Nulos:** Filas con al menos un valor faltante.
- **Duplicados:** Filas que se repiten.
- **Valores atípicos:** Filas con outliers según el método IQR (rango intercuartílico).

Cada aspecto se visualiza con un gráfico de dona (rojo/amarillo = problema, azul = sin problema).

5. Crear columna derivada

En el expander **Crear columna derivada (operación aritmética)** puedes:

1. Elegir la **primera columna** numérica.
2. Elegir la **segunda columna** numérica.
3. Seleccionar la **operación:** Suma, Resta, Multiplicación o División.
4. Asignar un **nombre** a la nueva columna.
5. Pulsar **Crear columna**.

Las columnas derivadas se mantienen entre recargas. Usa **Eliminar todas las columnas derivadas** para restablecer.

Pestaña 2: Visualización Dinámica (EDA)

Requiere haber ejecutado el procesamiento en la pestaña ETL.

Filtros globales

En **Configurar filtros** puedes restringir los datos usados en todas las visualizaciones:

- **Rango de fechas** — Si existe una columna de fecha, filtra por intervalo.
- **Categorías** — Filtra por valores de una columna categórica.
- **Slider numérico** — Filtra por rango de una columna numérica.

Análisis Univariado

Sección	Gráficos	Resumen con IA (si hay API Key)
Columnas numéricas	Histograma + Boxplot	“Resumen con IA (histograma y boxplot)”
Columnas categóricas	Gráfico de barras (conteo o %)	“Resumen con IA (gráfico de barras)”
Columnas booleanas	Gráfico de barras	“Resumen con IA (gráfico booleano)”

Análisis Bivariado

- **Correlaciones:** Heatmap de correlación entre columnas numéricas.
- **Evolución temporal:** Gráfico de líneas o área si hay columna de fecha (conteo, media, suma).
- **Numérico vs categórico:** Boxplot por categoría.
- **Numérico vs numérico:** Gráfico de dispersión (scatter).

Reporte

- Métricas: filas, columnas, % valores faltantes.
 - **Resumen estadístico** (`df.describe`) de columnas numéricas.
 - **Moda** por columna categórica/booleana.
 - **Descargar dataset filtrado** en formato CSV.
-

Pestaña 3: Insights de IA

Orientada a análisis guiado por IA. Requiere **API Key de Groq** configurada en la barra lateral.

Análisis principal

- **Generar Insights con IA** — Interpreta el resumen estadístico del dataset y genera:
 - **Tendencias** — Patrones en las variables.
 - **Riesgos** — Posibles problemas o sesgos.
 - **Oportunidades** — Mejoras o acciones recomendables.

Correlaciones

- Heatmap de correlación.
- **Generar notas sobre correlaciones** — Comentarios sobre correlaciones fuertes, multicolinealidad y pares de variables relevantes.

Valores faltantes e imputación

Si hay columnas con valores faltantes:

- **Sugerir método de imputación** — Recomendaciones por tipo de columna (media, mediana, moda, etc.).

Chat

En la columna derecha puedes hacer preguntas sobre el dataset. El asistente usa el contexto de los datos cargados para responder. Las respuestas se muestran en español.

Temas y personalización

La aplicación incluye un tema con colores institucionales (azul #004B85, amarillo #FFB903). Puedes alternar entre **modo claro** y **modo oscuro** en el menú de configuración (ícono de engranaje o hamburguesa en la esquina superior derecha).

Preguntas frecuentes

¿Por qué no veo los botones “Resumen con IA”?

Debes configurar la API Key de Groq en la barra lateral. Sin ella, las funcionalidades de IA no están disponibles.

¿Qué hago si mi archivo tiene números con comas (ej. 1,234.56)?

La aplicación detecta automáticamente formatos numéricos localizados y los convierte correctamente.

¿Los datos se envían a servidores externos?

Solo si usas la API de Groq: el resumen estadístico y el contexto del dataset se envían para generar los insights. Los datos originales no se almacenan en Groq.

¿Puedo usar datasets de cualquier tamaño?

Sí, pero datasets muy grandes pueden tardar más en procesarse. Los gráficos interactivos (Plotly) siguen funcionando bien con muchos puntos.

Panel Inteligente — Universidad EAFIT · Maestría en Ciencia de Datos