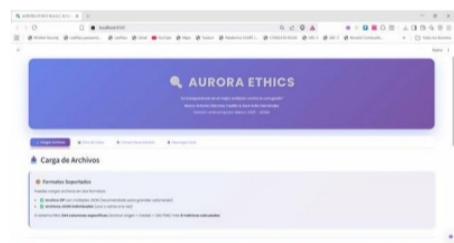


MANUAL DE USUARIO

AURORA ETHICS

Sistema de Análisis y Detección de
Irregularidades en Declaraciones Patrimoniales



“La transparencia es el mejor antídoto contra la corrupción”

Desarrollado por:

Marco Antonio Sánchez Cedillo (M.A.S.C.)
Saúl Ávila Hernández (S.A.H.)

January 25, 2026

Contents

1	Introducción	4
1.1	Origen del Proyecto	4
1.1.1	La Ley de Benford	4
1.1.2	Análisis de la Ley de Benford en el Proyecto	5
1.1.3	Ética, Transparencia y Generación de Valor	5
2	Parámetros Base del Sistema	6
2.1	Fundamentos del Análisis	6
2.1.1	Umbrales de Alerta Definidos	6
2.1.2	Identificación de Duplicados	6
2.1.3	Análisis de Máximos y Mínimos	7
2.1.4	Métricas Patrimoniales Calculadas	7
2.1.5	Sistema de Conteo sin Valores Cero	7
2.1.6	Cálculo de Alertas Individuales	8
2.1.7	Clasificación de Nivel de Riesgo	8
2.1.8	Variables Primarias y Secundarias	8
2.1.9	Normalización de Estados	9
3	Requerimientos Técnicos e Instalación	10
3.1	Instalación de Python	10
3.1.1	Paso 1: Descargar Python	10
3.2	Instalación de Visual Studio Code	10
3.2.1	Paso 2: Descargar Visual Studio Code	11
3.3	Instalación de Librerías	12
3.3.1	Paso 3: Instalar Librerías Necesarias	12
3.4	Descarga de Datos	12
3.4.1	Paso 4: Obtener Datos de la PDN	12
4	Ejecución del Programa	13
4.1	Inicio de la Aplicación	13
4.1.1	Paso 5: Ejecutar AURORA ETHICS	13
4.2	Carga de Datos JSON	14
4.2.1	Paso 6: Cargar Archivo JSON	14
4.3	Campo Personalizado	15
4.3.1	Paso 7: Agregar Campo Personalizado (Estado)	16
4.4	Descarga de Excel	16
4.4.1	Paso 8: Descargar Dataset Procesado	17
4.5	Carga Final y Análisis	19

4.5.1	Paso 9: Cargar Archivo Procesado	19
5	Secciones del Sistema	20
5.1	Pantalla Inicial	20
5.2	EDA - Análisis Exploratorio de Datos	21
5.2.1	Sección de Valores Únicos	22
5.2.2	Análisis de Valores Nulos	23
5.2.3	Análisis de Distribuciones	23
5.2.4	Matriz de Correlaciones	24
5.2.5	Top de Cargos con Mayores Ingresos	24
5.3	Visualización Global	25
5.4	Personalizados	26
5.5	Machine Learning / Geointeligencia	26
5.5.1	Mapa de Riesgo Geográfico	27
5.5.2	Análisis por Estado	28
5.5.3	Métricas Comparativas	28
5.6	Evolución Patrimonial (CORE del Proyecto)	29
5.6.1	Tabla Principal	29
5.6.2	Buscador Inteligente	32
5.6.3	Top 20 Casos de Alto Riesgo	33
5.6.4	Reporte de Métricas	34
5.7	Nepotismo	34
5.7.1	Red de Relaciones	35
5.7.2	Estadísticas del Grafo	35
5.7.3	Distribución por Tipo de Coincidencia	36
5.7.4	Top 20 Casos de Similitud	36
5.7.5	Entes con Mayor Coincidencia	36
5.7.6	Tabla de Casos Sospechosos	37
5.8	Sistema de Denuncias Ciudadanas	37
5.8.1	Registrar Nueva Denuncia	37
5.8.2	Historial de Denuncias	38
5.8.3	WordCloud de Denuncias	38
5.9	Reporte Ejecutivo	39
6	Escalabilidad: Implementación con MongoDB	40
6.1	Ejecución Local vs. Producción	40
6.2	¿Por Qué MongoDB?	40
6.3	Arquitectura Propuesta	40
6.4	Instalación de MongoDB	41
6.4.1	Paso 1: Descargar MongoDB Community Edition	41
6.4.2	Paso 2: Instalar MongoDB	41
6.4.3	Paso 3: Verificar la Instalación	41
6.4.4	Paso 4: Iniciar el Servicio MongoDB	41
6.5	Integración con AURORA ETHICS	42
6.5.1	Paso 5: Instalar PyMongo	42
6.5.2	Paso 6: Configurar la Conexión	42
6.5.3	Paso 7: Insertar Datos en MongoDB	43
6.5.4	Paso 8: Consultar Datos desde MongoDB	43
6.6	Ventajas de la Implementación con MongoDB	43

6.7 Seguridad y Autenticación	44
6.8 MongoDB Atlas (Nube)	44
Agradecimientos	45

Chapter 1

Introducción

1.1 Origen del Proyecto

Nuestro interés en el desarrollo de AURORA ETHICS nació de la experiencia que hemos acumulado en el análisis de información y datos a lo largo de nuestra trayectoria profesional. En nuestro *background* hemos aprendido a utilizar diferentes enfoques y metodologías analíticas, entre ellas la técnica de la Ley de Benford, herramienta fundamental que ha demostrado ser altamente efectiva en la detección de anomalías en conjuntos de datos financieros.

1.1.1 La Ley de Benford

La **Ley de Benford**, también conocida como la *Ley del Primer Dígito*, es un principio matemático que describe la distribución de los primeros dígitos en conjuntos de datos numéricos que ocurren naturalmente. Descubierta por el astrónomo Simon Newcomb en 1881 y posteriormente popularizada por el físico Frank Benford en 1938, esta ley establece que en muchos conjuntos de datos del mundo real, el dígito 1 aparece como primer dígito aproximadamente el 30.1% de las veces, mientras que el 9 solo aparece cerca del 4.6%.

¿Por qué es relevante en AURORA ETHICS?

En el contexto de nuestro proyecto de análisis de declaraciones patrimoniales, la Ley de Benford cobra especial relevancia porque:

- Los datos financieros naturales (ingresos, patrimonio, bienes) tienden a seguir esta distribución
- Las desviaciones significativas de esta ley pueden indicar manipulación o fabricación de datos
- Permite identificar patrones sospechosos que merecen una investigación más profunda
- Es una herramienta complementaria robusta para validar la autenticidad de las declaraciones

1.1.2 Análisis de la Ley de Benford en el Proyecto

A continuación se muestra la gráfica de la Ley de Benford aplicada a los ingresos de los servidores públicos analizados:

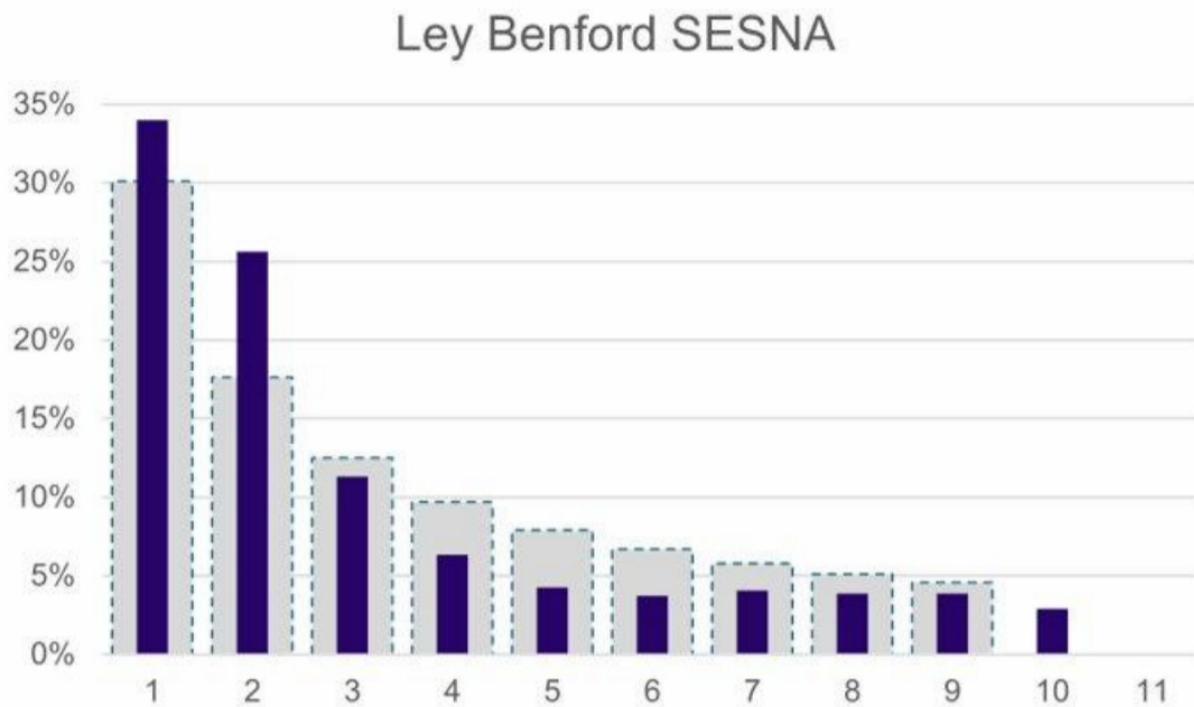


Figure 1.1: Distribución de primeros dígitos en ingresos declarados vs. Ley de Benford

Interpretación de resultados:

El análisis de la gráfica revela que la distribución de los primeros dígitos en los ingresos declarados de los servidores públicos muestra una adherencia general a la Ley de Benford, con algunas desviaciones notables. En particular, se observa:

- El dígito 1 aparece con una frecuencia cercana a lo esperado (aproximadamente 30%)
- Existen pequeñas variaciones en los dígitos intermedios (4, 5, 6) que podrían indicar efectos de redondeo o tabuladores salariales estandarizados
- No se detectan desviaciones extremas que sugieran fabricación masiva de datos
- La distribución global valida la autenticidad general del conjunto de datos

1.1.3 Ética, Transparencia y Generación de Valor

Estamos profundamente interesados en temas de **ética** y **transparencia** en el servicio público. Creemos firmemente que la tecnología y el análisis de datos son herramientas poderosas para generar valor social, fortalecer la rendición de cuentas y contribuir a la construcción de instituciones más confiables y eficientes.

AURORA ETHICS representa nuestro compromiso con estos valores, proporcionando una plataforma robusta, científica y accesible para el análisis de declaraciones patrimoniales y la detección de posibles irregularidades que requieran mayor escrutinio.

Chapter 2

Parámetros Base del Sistema

2.1 Fundamentos del Análisis

AURORA ETHICS está construido sobre una base sólida de parámetros y métricas que permiten identificar alertas de corrupción de manera sistemática y fundamentada. A continuación se detallan todos los parámetros utilizados en el sistema.

2.1.1 Umbrales de Alerta Definidos

El sistema utiliza tres umbrales principales para la detección de anomalías:

Umbrales Críticos
<ul style="list-style-type: none">UMBRAL_ALERTA_INCREMENTO = 100.0% Porcentaje de incremento patrimonial que se considera inusual. Un incremento superior al 100% entre declaraciones genera una alerta de alto riesgo.UMBRAL_RATIO BIENES_INGRESOS = 3.0 Ratio entre el valor de bienes inmuebles y los ingresos anuales. Si los bienes superan 3 veces los ingresos, se genera alerta.UMBRAL_RATIO_ADEUDOS_INGRESOS = 1.5 Ratio entre adeudos totales e ingresos anuales. Adeudos superiores a 1.5 veces los ingresos generan alerta de riesgo financiero.

2.1.2 Identificación de Duplicados

El sistema realiza un análisis exhaustivo para identificar duplicados en:

- Bienes inmuebles:** Detecta propiedades declaradas repetidamente con el mismo valor de adquisición
- Vehículos:** Identifica automóviles duplicados basándose en valor y tipo
- Inversiones:** Filtra inversiones reportadas múltiples veces

- **Adeudos:** Elimina doble contabilización de deudas

Esta depuración es crucial para evitar sobreestimar el patrimonio o las obligaciones de un servidor público.

2.1.3 Análisis de Máximos y Mínimos

Para cada servidor público, el sistema calcula:

- **Máximo ingreso del servidor público (SP):** Identifica el mayor ingreso declarado por concepto de salario público
- **Mínimo ingreso del SP:** Registra el menor ingreso oficial declarado
- **Máximo otros ingresos (OI):** Captura el mayor monto reportado en otros ingresos
- **Mínimo otros ingresos:** Documenta el menor valor de ingresos adicionales
- **Máximo ingreso total:** Suma el punto más alto de todos los ingresos
- **Mínimo ingreso total:** Identifica el periodo con menores ingresos

2.1.4 Métricas Patrimoniales Calculadas

El sistema genera automáticamente las siguientes métricas para cada servidor público:

1. **Patrimonio_Total:** Suma de bienes inmuebles + vehículos + bienes muebles + inversiones - adeudos
2. **Ratio_Patrimonio_Ingreso:** Patrimonio total dividido entre ingresos anuales promedio
3. **Delta_Patrimonial_Anual:** Cambio absoluto en patrimonio entre declaraciones
4. **Incremento_Porcentual_Patrimonio:** Porcentaje de crecimiento patrimonial
5. **Años_en_Cargo:** Tiempo transcurrido desde la toma de posesión del cargo
6. **Tiene_Conflicto_Interes:** Indicador binario de participación en empresas
7. **Numero_Bienes_Totales:** Conteo de todos los activos declarados
8. **Ratio_Deuda_Activos:** Proporción de deudas respecto a activos totales

2.1.5 Sistema de Conteo sin Valores Cero

Para evitar distorsiones estadísticas, el sistema implementa conteos especiales:

- **Decla_sin0SP:** Número de declaraciones con ingreso del servidor público mayor a cero
- **Decla_sin0OI:** Declaraciones con otros ingresos distintos de cero
- **Decla_sin0TI:** Total de declaraciones con ingresos totales no nulos

Estos conteos permiten calcular promedios más precisos, excluyendo períodos donde legítimamente no hubo ingresos de cierta categoría.

2.1.6 Cálculo de Alertas Individuales

El sistema genera alertas específicas basadas en múltiples indicadores:

Tipos de Alertas

1. **INCREMENTO_INUSUAL**: Incremento patrimonial 100%
2. **BIENES_ALTOS**: Ratio bienes/ingresos 3.0x
3. **ADEUDOS_ELEVADOS**: Ratio adeudos/ingresos 1.5x
4. **OTROS_INGRESOS_ALTOS**: Otros ingresos 2.0x ingreso del SP
5. **INVERSIONES_ALTAS**: Inversiones 1.5x ingresos totales
6. **VEHICULOS_COSTOSOS**: Promedio de vehículo 2x ingreso anual

2.1.7 Clasificación de Nivel de Riesgo

El sistema asigna puntos de riesgo y clasifica a cada servidor público:

Nivel de Riesgo	Puntos	Criterio
ALTO	7+	Múltiples alertas graves
MEDIO	4-6	Alertas moderadas
BAJO	0-3	Sin alertas significativas

Table 2.1: Clasificación de niveles de riesgo

Los puntos se acumulan de la siguiente manera:

- Incremento inusual: +3 puntos
- Bienes desproporcionados: +2 puntos (+2 adicionales si 6.0x)
- Adeudos elevados: +2 puntos
- Otros ingresos altos: +2 puntos (+2 adicionales si 5.0x)
- Inversiones no justificadas: +1 punto
- Valores cero sospechosos: +1 punto

2.1.8 Variables Primarias y Secundarias

Variable Primaria: Ingreso del Servidor Público (SP)

Esta es la variable fundamental que sirve como base de comparación:

- Se obtiene del campo `remuneracionAnualCargoPublico`
- Si está en cero, se anualiza desde `remuneracionMensualCargoPublico` x 12
- Sirve como referencia para calcular todos los ratios

Variables Secundarias que Generan Alertas:

1. **Otros Ingresos:** Actividades independientes, arrendamientos, etc.
2. **Bienes Inmuebles:** Casas, terrenos, departamentos
3. **Vehículos:** Automóviles, motocicletas
4. **Inversiones:** Cuentas bancarias, inversiones, ahorros
5. **Adeudos:** Hipotecas, préstamos, tarjetas de crédito
6. **Bienes Muebles:** Joyas, obras de arte, menaje

Cada una de estas variables se compara contra el ingreso del SP para determinar si existe desproporcionalidad que amerite una alerta.

2.1.9 Normalización de Estados

El sistema normaliza los nombres de estados a una lista válida de 18 entidades:²

- Aguascalientes
- Baja California
- Baja California Sur
- Chiapas
- Chihuahua
- Ciudad de México
- Estado de México
- Guanajuato
- Guerrero
- Jalisco
- Michoacán
- Morelos
- Nayarit
- Quintana Roo
- San Luis Potosí
- Sinaloa
- Tabasco
- Zacatecas

Esta normalización permite realizar análisis geográficos consistentes y comparar entidades de manera uniforme.

Chapter 3

Requerimientos Técnicos e Instalación

3.1 Instalación de Python

3.1.1 Paso 1: Descargar Python

Acceder a la página oficial de Python: <https://www.python.org/downloads/>



Figure 3.1: Página oficial de descarga de Python

Importante: Durante la instalación, es crucial activar la opción “**Add Python to PATH**” para que las librerías puedan instalarse posteriormente desde el CMD usando pip install.

Descargar la versión actual o que corresponda a su sistema operativo. Se recomienda Python 3.9 o superior.

3.2 Instalación de Visual Studio Code

3.2.1 Paso 2: Descargar Visual Studio Code

Acceder a: <https://code.visualstudio.com/download>

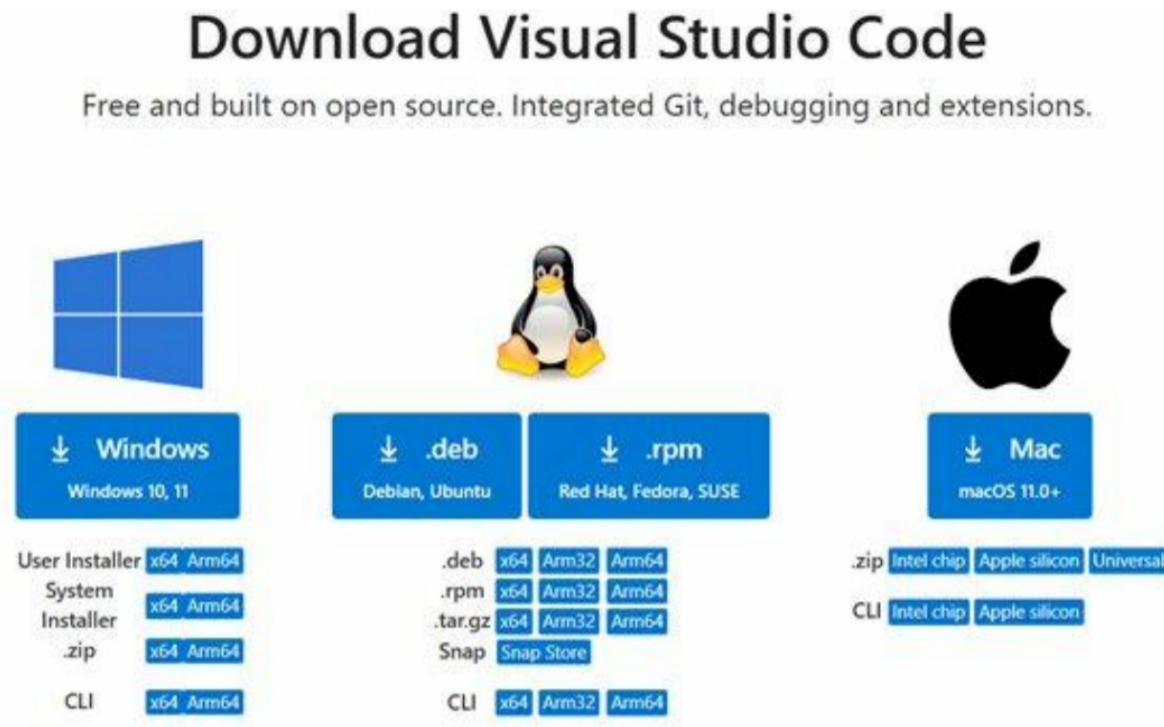


Figure 3.2: Página de descarga de Visual Studio Code

Posteriormente, dentro de Visual Studio Code, activar los complementos de Python para poder ejecutar todos los programas.

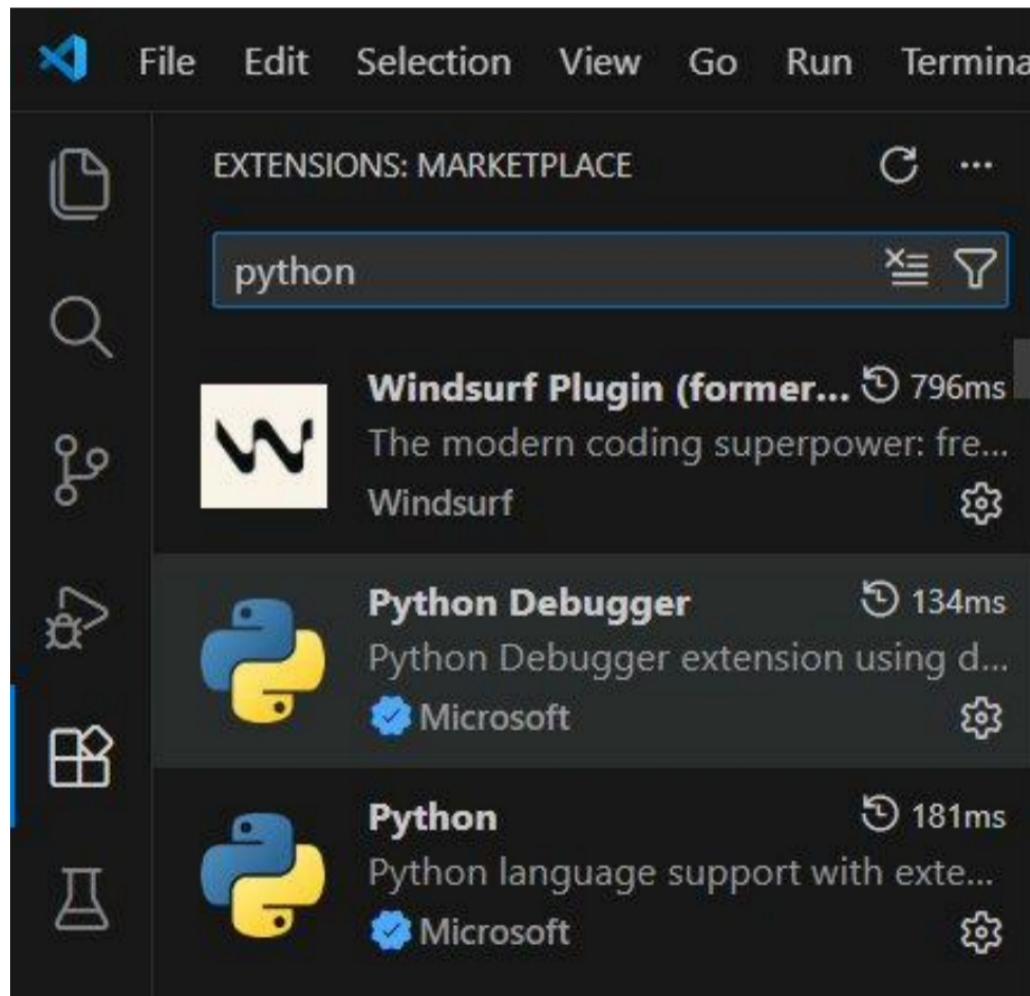


Figure 3.3: Instalación de complementos de Python en VS Code

3.3 Instalación de Librerías

3.3.1 Paso 3: Instalar Librerías Necesarias

Una vez que se tienen los requisitos previos instalados, abrir el CMD (Símbolo del sistema) en Windows y ejecutar el siguiente comando completo en una sola línea:

```
pip install streamlit pandas numpy matplotlib seaborn plotly  
scipy scikit-learn fuzzywuzzy networkx openpyxl wordcloud chardet  
python-Levenshtein
```

Este comando instalará todas las librerías necesarias para el correcto funcionamiento de AURORA ETHICS:

- **streamlit**: Framework para la interfaz web
- **pandas**: Manipulación y análisis de datos
- **numpy**: Operaciones numéricas
- **matplotlib**: Visualización de datos básica
- **seaborn**: Visualización estadística avanzada
- **plotly**: Gráficos interactivos
- **scipy**: Algoritmos científicos y estadísticos
- **scikit-learn**: Machine learning y análisis predictivo
- **fuzzywuzzy**: Comparación difusa de textos
- **networkx**: Análisis de redes y grafos
- **openpyxl**: Manejo de archivos Excel
- **wordcloud**: Generación de nubes de palabras
- **chardet**: Detección automática de codificación
- **python-Levenshtein**: Optimización para fuzzywuzzy

3.4 Descarga de Datos

3.4.1 Paso 4: Obtener Datos de la PDN

Ingresar a la Plataforma Digital Nacional: <https://www.plataformadigitalnacional.org/>

Dirigirse al Sistema 1 - Declaraciones: <https://www.plataformadigitalnacional.org/declaraciones>

Descargar la base de datos en formato JSON de las declaraciones patrimoniales. El archivo descargado vendrá comprimido en formato ZIP.

Extraer la información de la carpeta ZIP en una ubicación conocida de su computadora.

Chapter 4

Ejecución del Programa

4.1 Inicio de la Aplicación

4.1.1 Paso 5: Ejecutar AURORA ETHICS

Para proceder a ejecutar el programa que funciona de manera local:

1. Abrir el navegador de Windows y buscar “**Windows PowerShell**”
2. Colocar el comando `cd`, seguido de un espacio, y luego la dirección donde está ubicado el programa `AURORA_ETHICS.py`
Ejemplo: `cd C:\Users\Usuario\Documentos\AURORA`
3. Presionar Enter
4. Ahora ejecutar el siguiente comando:

```
streamlit run AURORA_ETHICS.py
```

De manera automática se abrirá la aplicación en el navegador predeterminado.

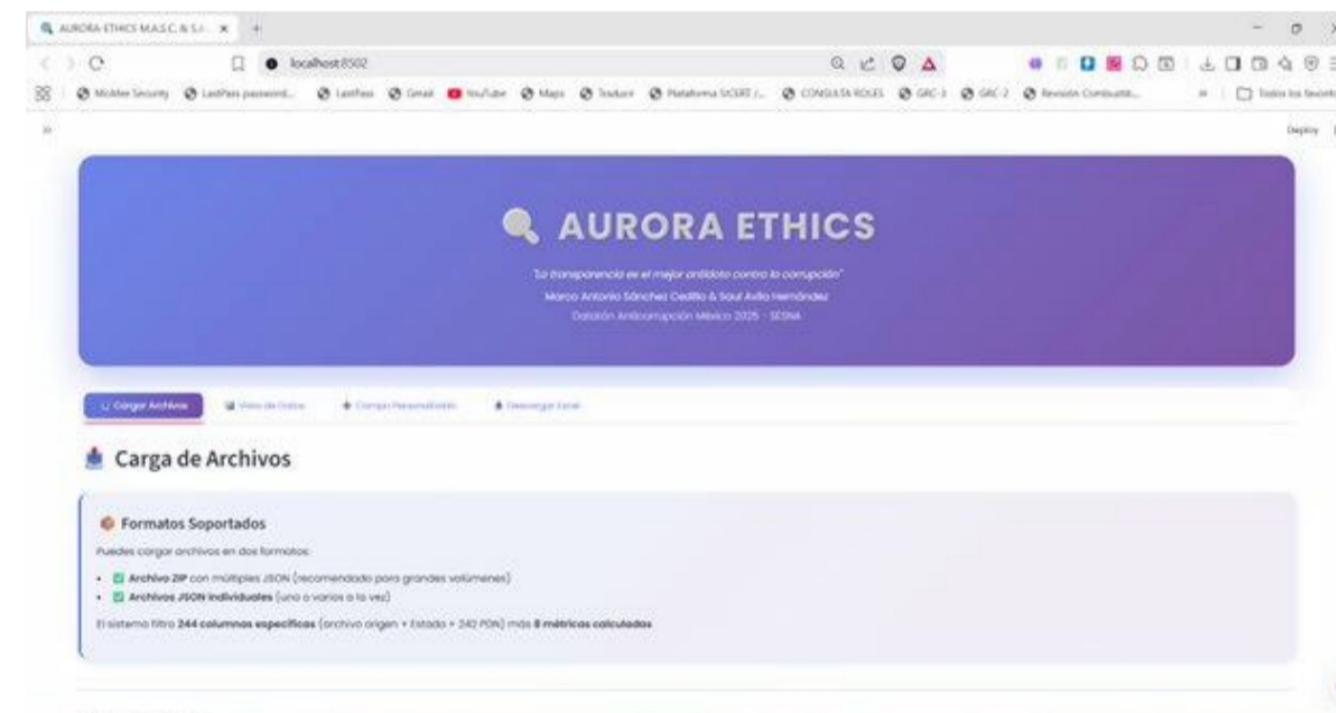


Figure 4.1: Pantalla principal de AURORA ETHICS

4.2 Carga de Datos JSON

4.2.1 Paso 6: Cargar Archivo JSON

En la primera pestaña, seleccionar el botón “Cargar Archivo JSON”:

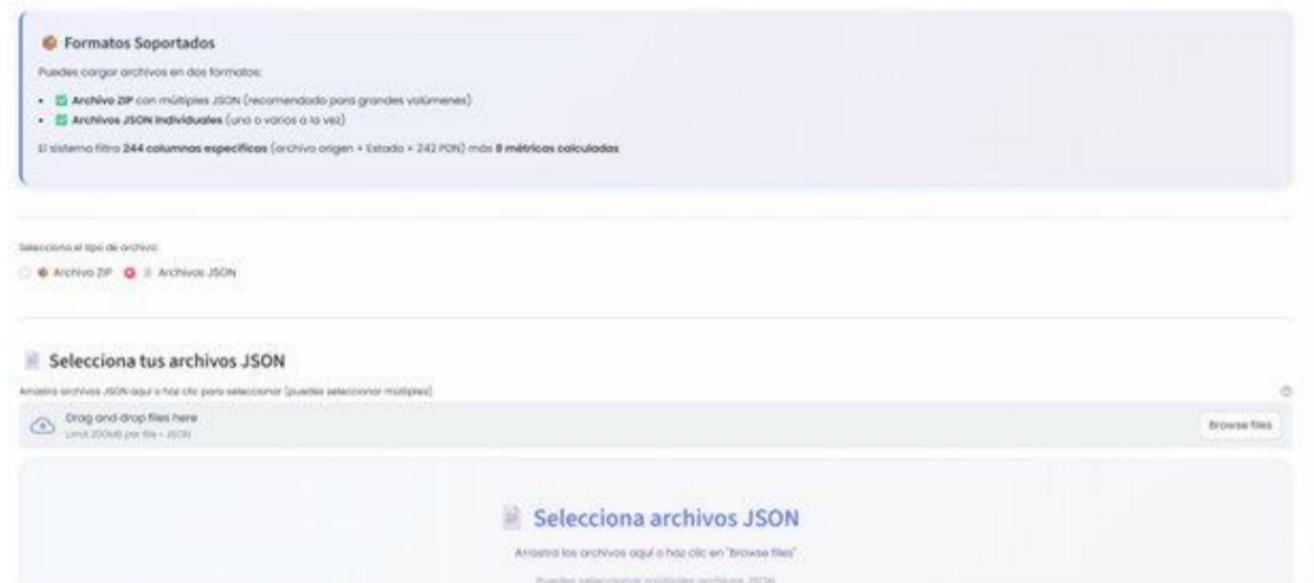


Figure 4.2: Botón para cargar archivos JSON

Después, presionar el botón “Cargar Archivo”:

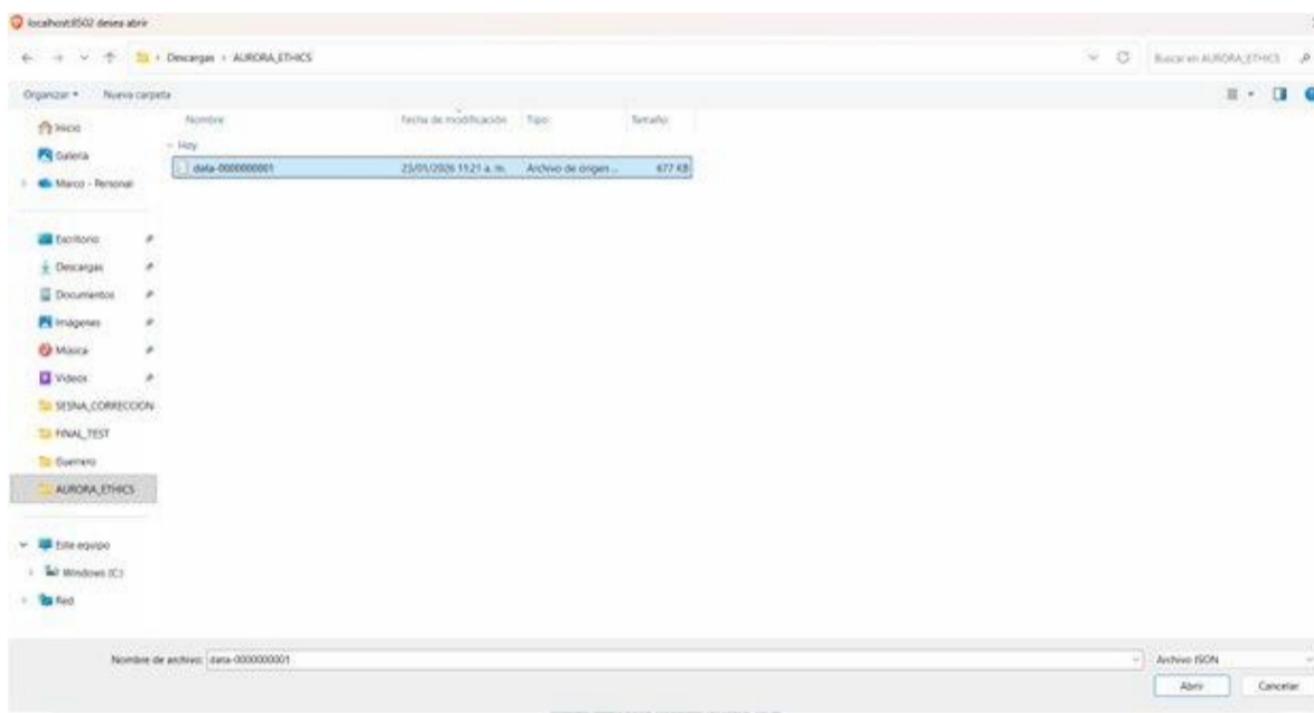


Figure 4.3: Selección de archivo JSON

Se debe seleccionar el archivo JSON descargado de la PDN y presionar “Abrir”:

Al cargar el archivo, se habilitará el botón “Procesar y Consolidar Datos JSON”.

Presionarlo:



Figure 4.4: Botón para procesar datos

Aparecerán globos de carga exitosa y una barra de progreso:

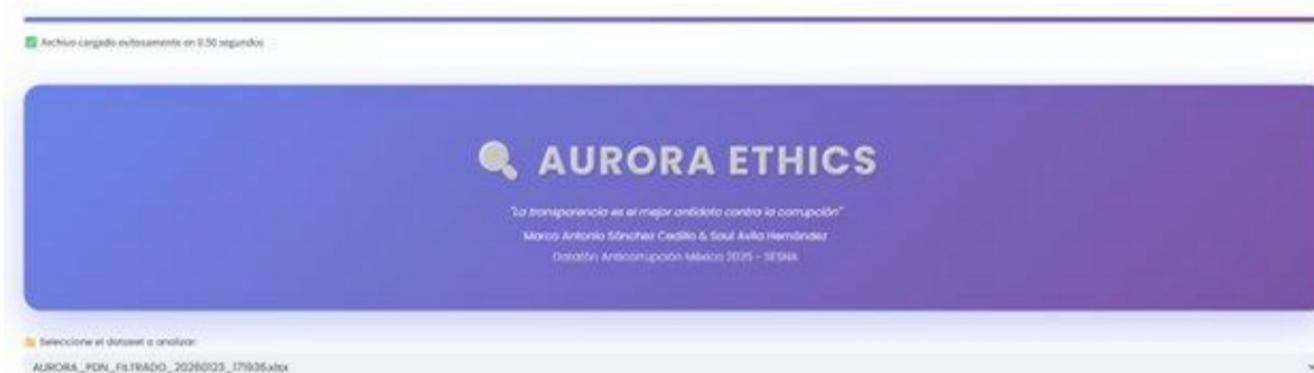


Figure 4.5: Barra de tiempo de carga

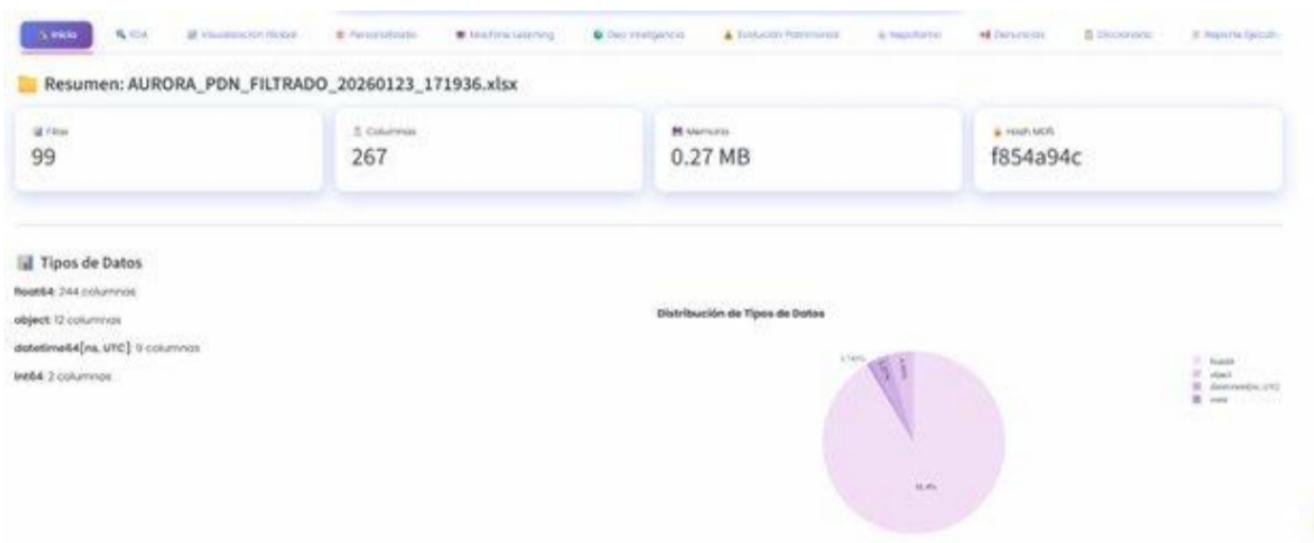


Figure 4.6: Resumen de información procesada

4.3 Campo Personalizado

4.3.1 Paso 7: Agregar Campo Personalizado (Estado)

Ir a la parte superior y seleccionar la pestaña “Campo Personalizado”:

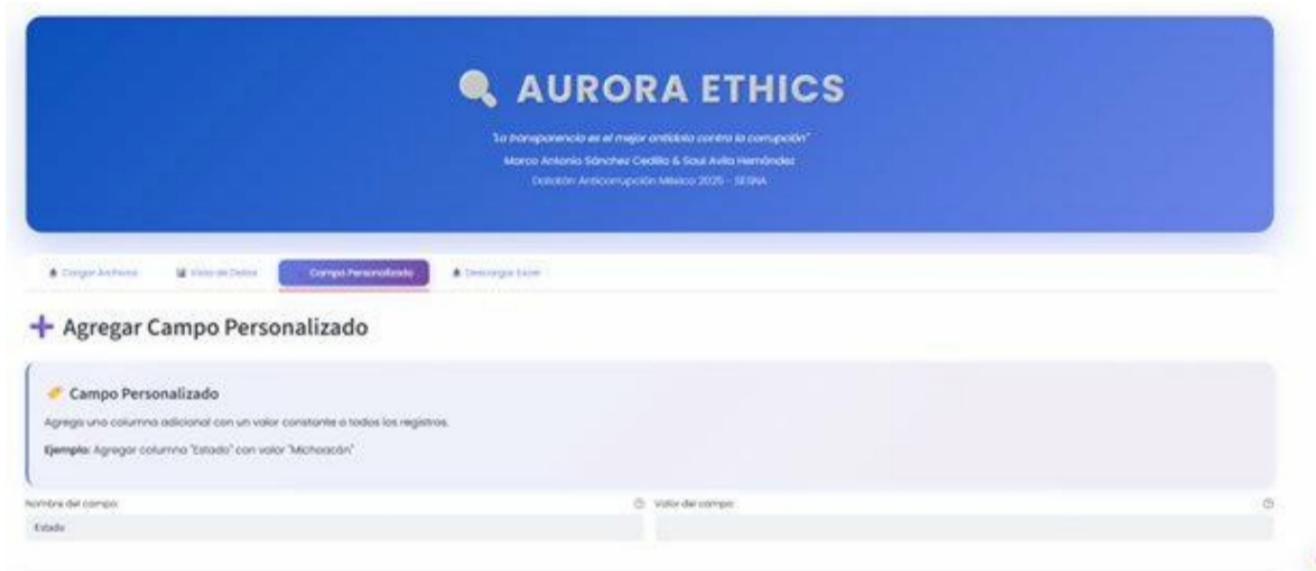


Figure 4.7: Pestaña de campo personalizado

Escribir el estado que se está leyendo (por ejemplo, “Guerrero”) y presionar el botón “+ Agregar Campo”:



Figure 4.8: Agregar campo personalizado

Se mostrará el mensaje de confirmación: “*Campo ‘Estado’ agregado exitosamente con valor ‘Guerrero’.*”

4.4 Descarga de Excel

4.4.1 Paso 8: Descargar Dataset Procesado

Ir a la pestaña “Descargar Excel”:

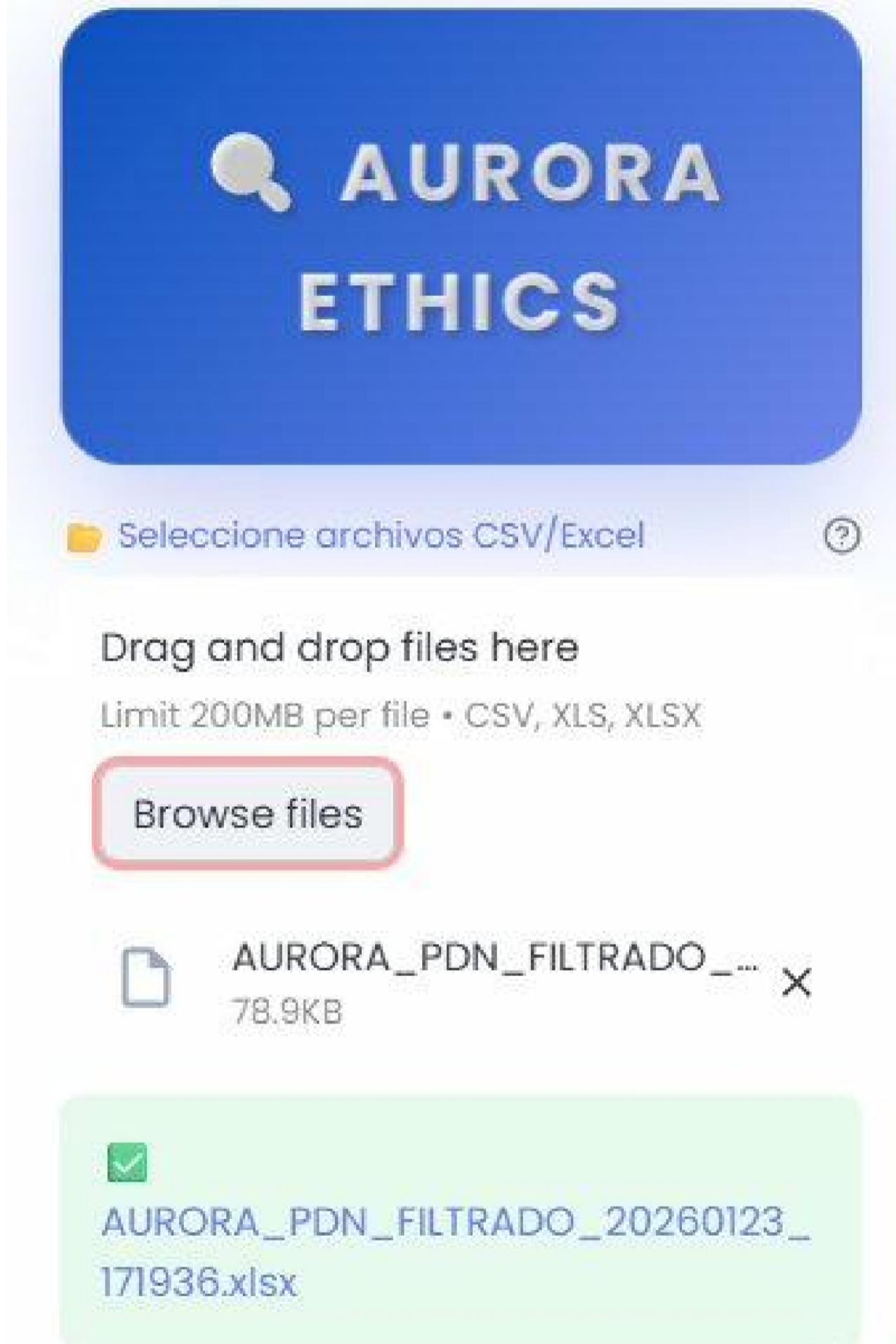


Figure 4.9: Pestaña de descarga de Excel

En esta sección ya se encuentra el dataset únicamente con las columnas seleccionadas. Bajar un poco la pantalla y presionar el botón “**Descargar Excel**”:



Figure 4.10: Botón para descargar Excel

Se guardará el archivo procesado con todas las métricas calculadas.

4.5 Carga Final y Análisis

4.5.1 Paso 9: Cargar Archivo Procesado

Abrir el panel izquierdo y cargar el archivo procesado (el archivo Excel descargado en el paso anterior):

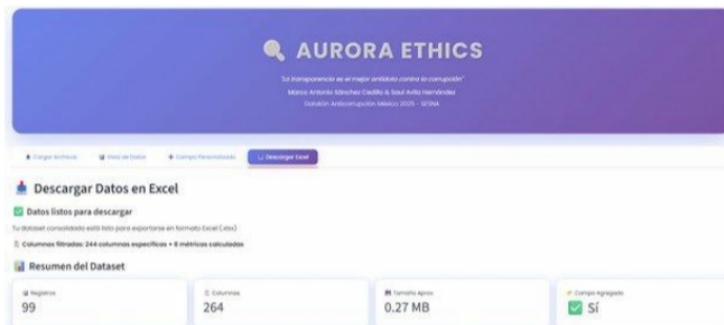


Figure 4.11: Panel lateral para cargar datos

Automáticamente se procesarán los datos y aparecerá la barra de tiempo de carga.

Ahora se debe ir hacia abajo hasta donde se encuentran los tabs de análisis, debajo del botón “**PROCESAR Y CONSOLIDAR JSON**”.

¡El sistema está listo para analizar! Toda la información procesada estará disponible en las diferentes pestañas.

Chapter 5

Secciones del Sistema

5.1 Pantalla Inicial

La pantalla inicial muestra información general del archivo cargado:

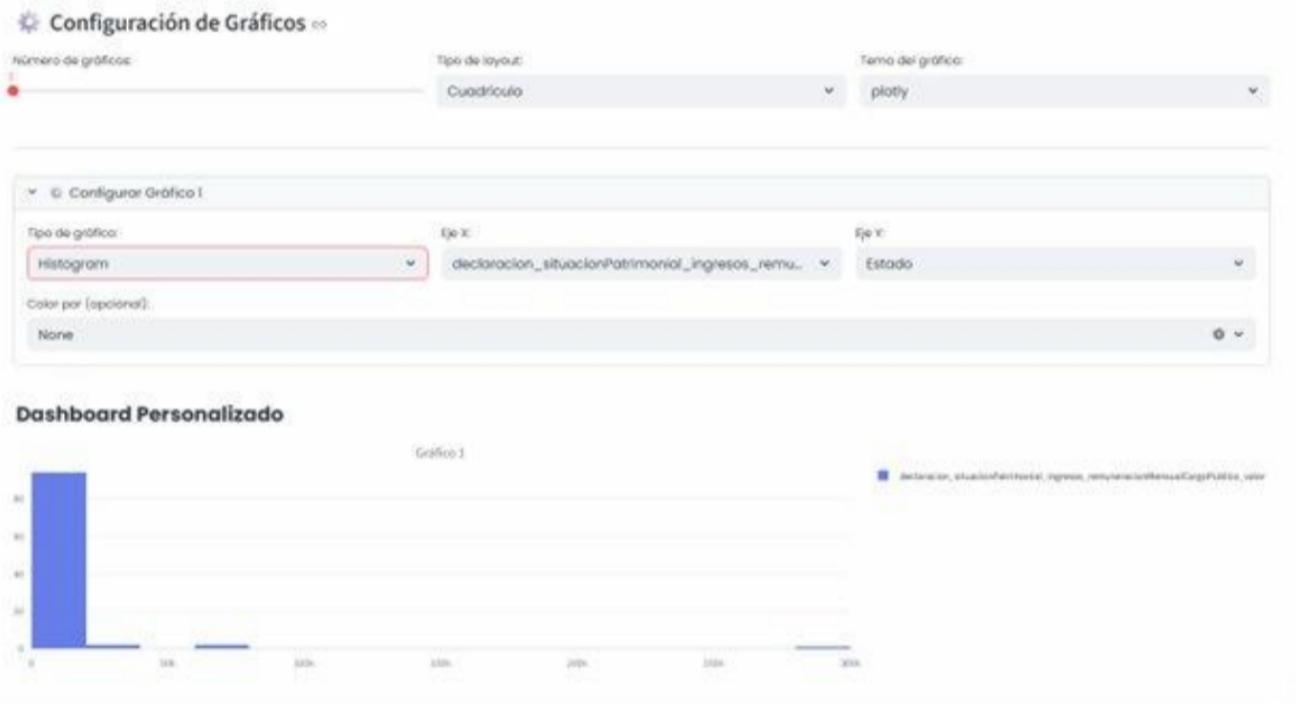


Figure 5.1: Pantalla inicial con información general

Se presenta un resumen ejecutivo con:

- **Filas:** Número total de registros
- **Columnas:** Cantidad de variables en el dataset
- **Memoria:** Tamaño aproximado en MB
- **Hash MD5:** Identificador único del archivo

Adicionalmente, se muestra la distribución de tipos de datos:

- **float64:** Columnas numéricas decimales
- **object:** Columnas de texto

- **datetime64[ns, UTC]**: Columnas de fecha/hora
- **int64**: Columnas numéricas enteras

Una gráfica visual presenta la distribución de tipos de datos en el conjunto.

5.2 EDA - Análisis Exploratorio de Datos

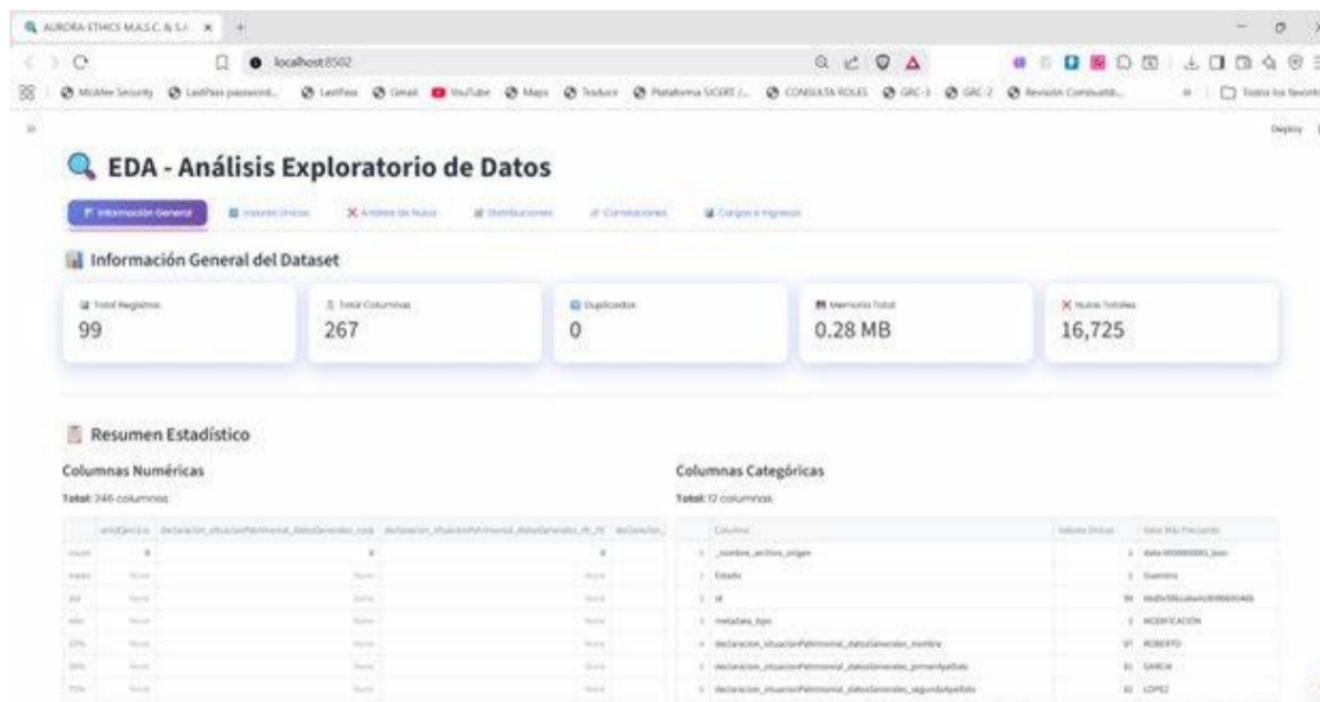


Figure 5.2: Sección de EDA - Análisis Estadístico

Esta sección contiene el análisis estadístico detallado de los datos, incluyendo:

- Tipos de datos por columna
- Valores más frecuentes (top values)
- Estadísticas descriptivas (media, mediana, desviación estándar)
- Distribuciones de frecuencia

5.2.1 Sección de Valores Únicos

Se puede seleccionar una columna específica y el sistema desplegará gráficos automáticos del dataset:

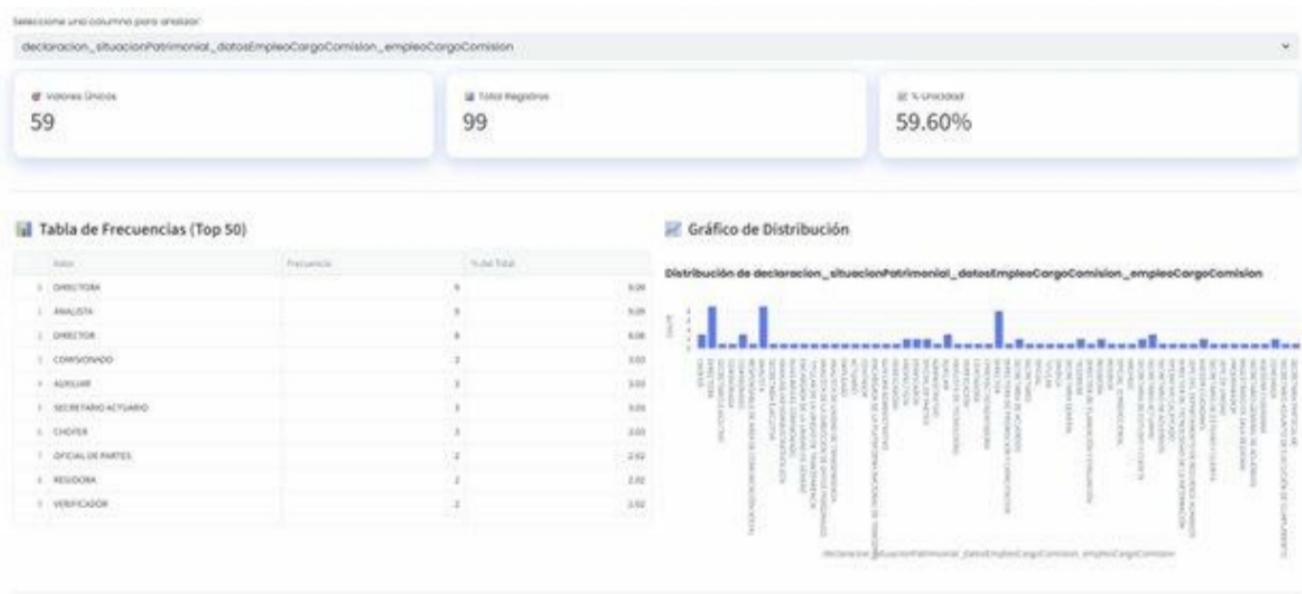


Figure 5.3: Visualización de valores únicos - Parte 1

Nube de Palabras: declaracion_situacionPatrimonial_datosEmpleoCargoComision_emploiCargoComision



Figure 5.4: Visualización de valores únicos - Parte 2

5.2.2 Análisis de Valores Nulos

El sistema muestra información detallada sobre columnas con datos nulos:

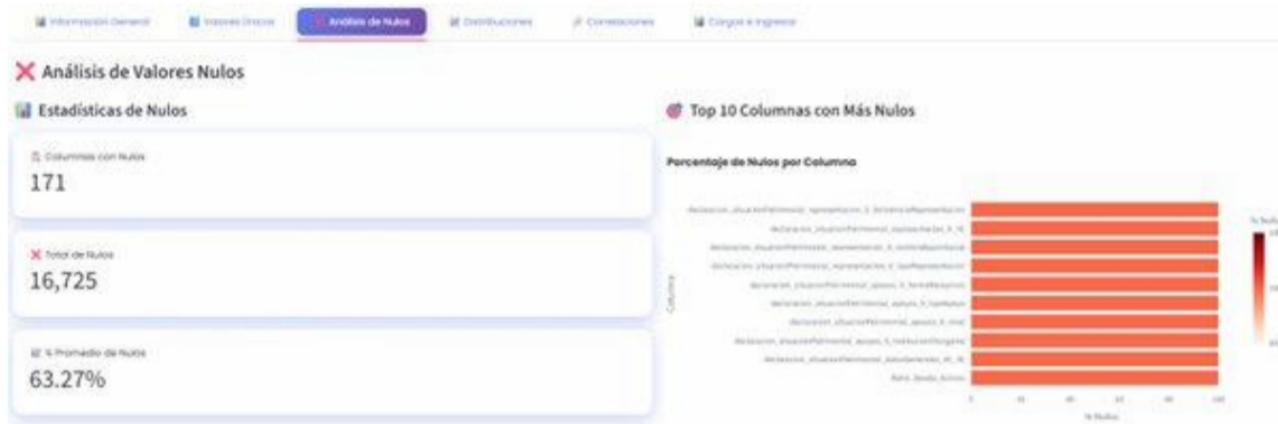


Figure 5.5: Análisis de valores nulos por columna

5.2.3 Análisis de Distribuciones

Se presenta un análisis estadístico de las distribuciones de las variables:

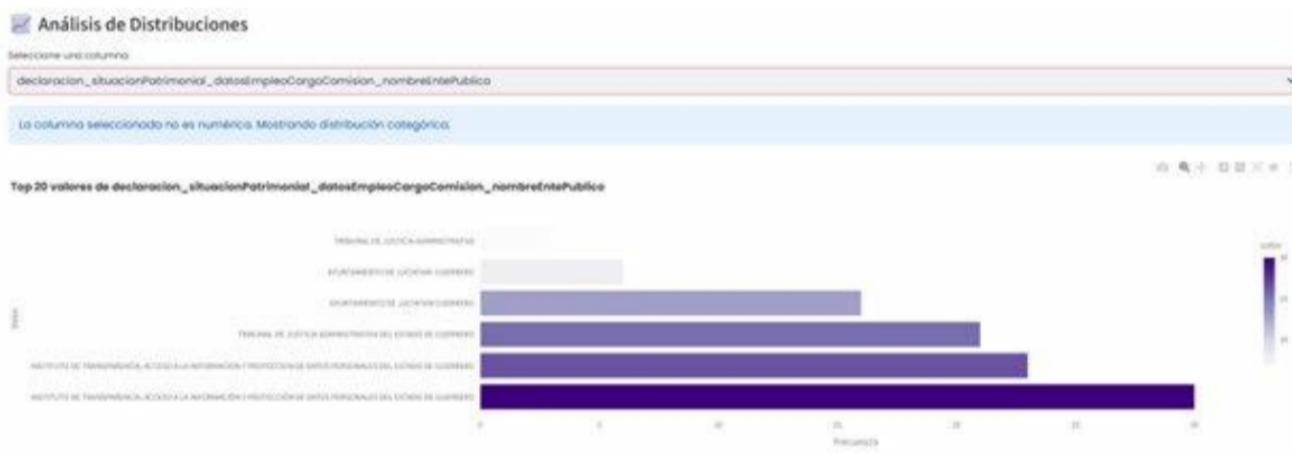


Figure 5.6: Análisis de distribuciones estadísticas

Al seleccionar una columna específica, el sistema genera histogramas y análisis de distribución automáticos.

5.2.4 Matriz de Correlaciones

Una de las herramientas más poderosas es la matriz de correlaciones entre variables:



Figure 5.7: Matriz de correlaciones entre variables

Esta matriz permite identificar relaciones entre diferentes métricas patrimoniales y detectar patrones sospechosos.

5.2.5 Top de Cargos con Mayores Ingresos

El sistema identifica y visualiza los cargos con mayores ingresos promedio:



Figure 5.8: Top de cargos con mayores ingresos promedio

5.3 Visualización Global

Dentro de la pestaña “Visualización Global” se pueden seleccionar variables para obtener gráficos generales de la base de datos:

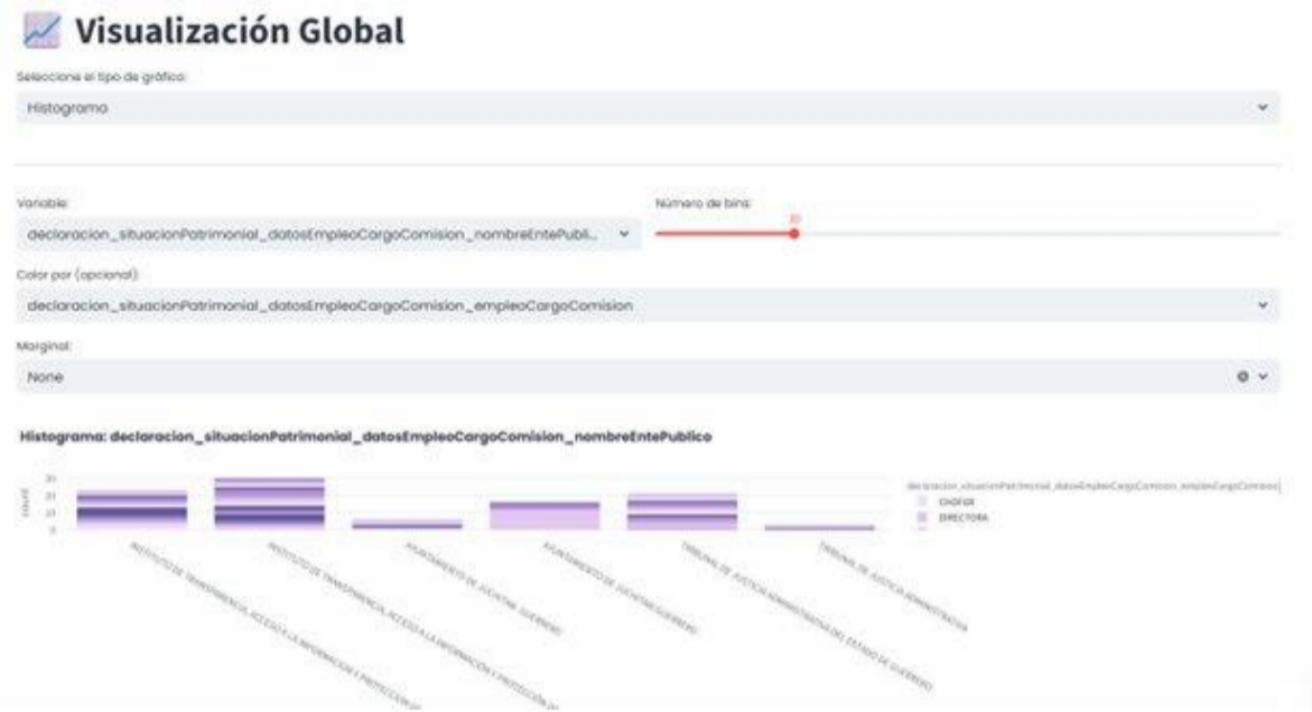


Figure 5.9: Visualización global de variables

Esta sección permite:

- Seleccionar múltiples variables simultáneamente
- Generar gráficos de barras, líneas, dispersión
- Comparar diferentes métricas visualmente
- Exportar visualizaciones

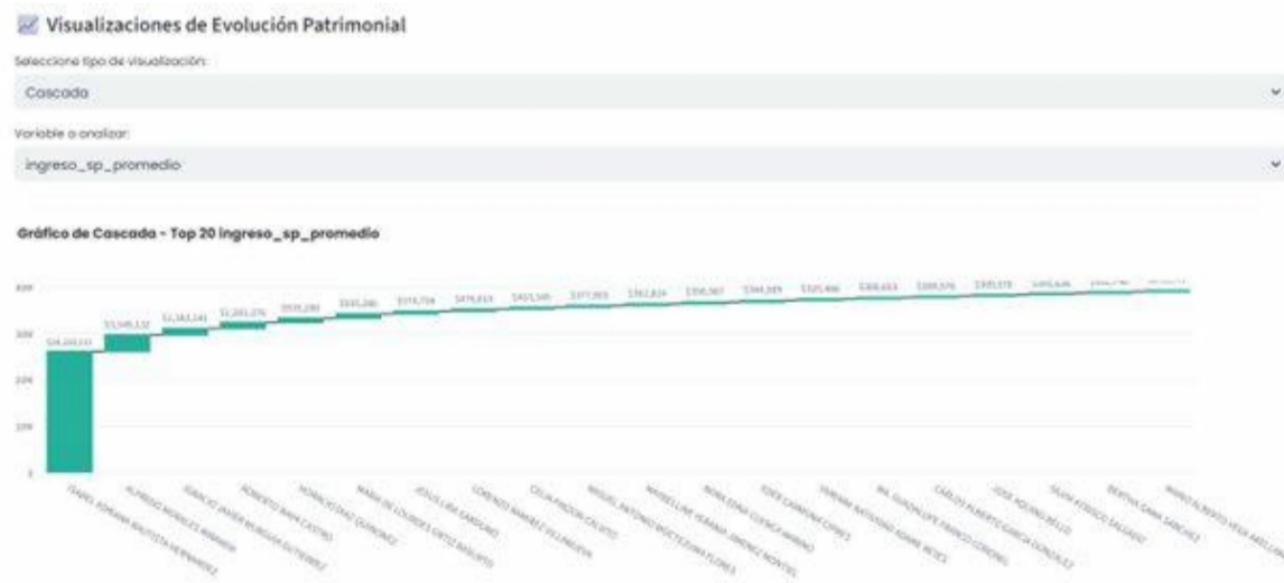


Figure 5.10: Visualización de ingresos promedio determinados

5.4 Personalizados

La sección de “Personalizados” permite seleccionar el número de gráficos deseados y las variables que se utilizarán para graficar:

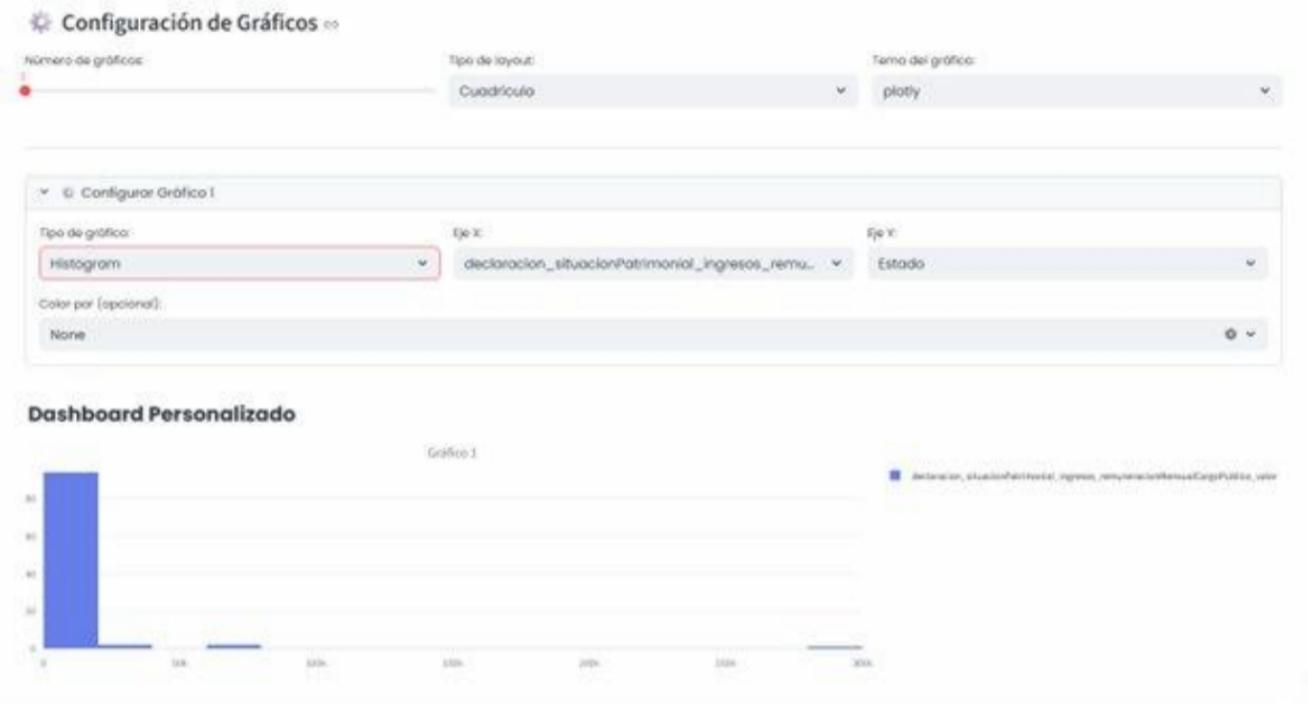


Figure 5.11: Gráficos personalizados

Esta funcionalidad ofrece máxima flexibilidad para:

- Crear dashboards personalizados
- Combinar métricas específicas
- Analizar relaciones particulares
- Generar reportes visuales ad-hoc

5.5 Machine Learning / Geointeligencia

5.5.1 Mapa de Riesgo Geográfico

La sección de “**Geointeligencia**” presenta tres subsecciones. La primera es el mapa de riesgo donde se visualizan los datos en el mapa de la República Mexicana:



Figure 5.12: Mapa de riesgo geográfico

El mapa indica si la alerta a nivel estado tiene riesgo **BAJO**, **MEDIO** o **ALTO**, utilizando un código de colores:

- **Verde:** Riesgo BAJO
- **Amarillo:** Riesgo MEDIO
- **Rojo:** Riesgo ALTO

5.5.2 Análisis por Estado

La segunda sección visualiza los servidores con mayor riesgo por estado:

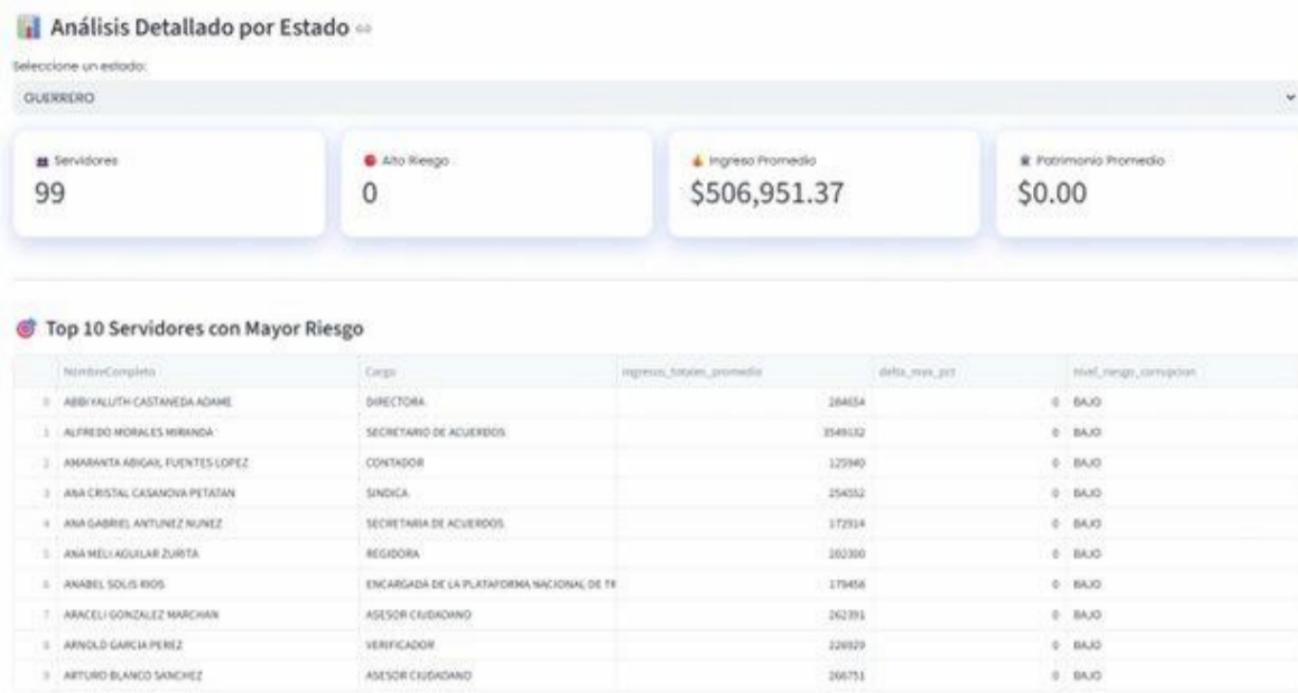


Figure 5.13: Servidores con mayor riesgo por estado

Esta vista permite identificar:

- Estados con mayor concentración de casos de riesgo
- Servidores individuales que requieren investigación prioritaria
- Patrones geográficos de irregularidades

5.5.3 Métricas Comparativas

La última sección presenta métricas comparativas de los ingresos promedio por estado:

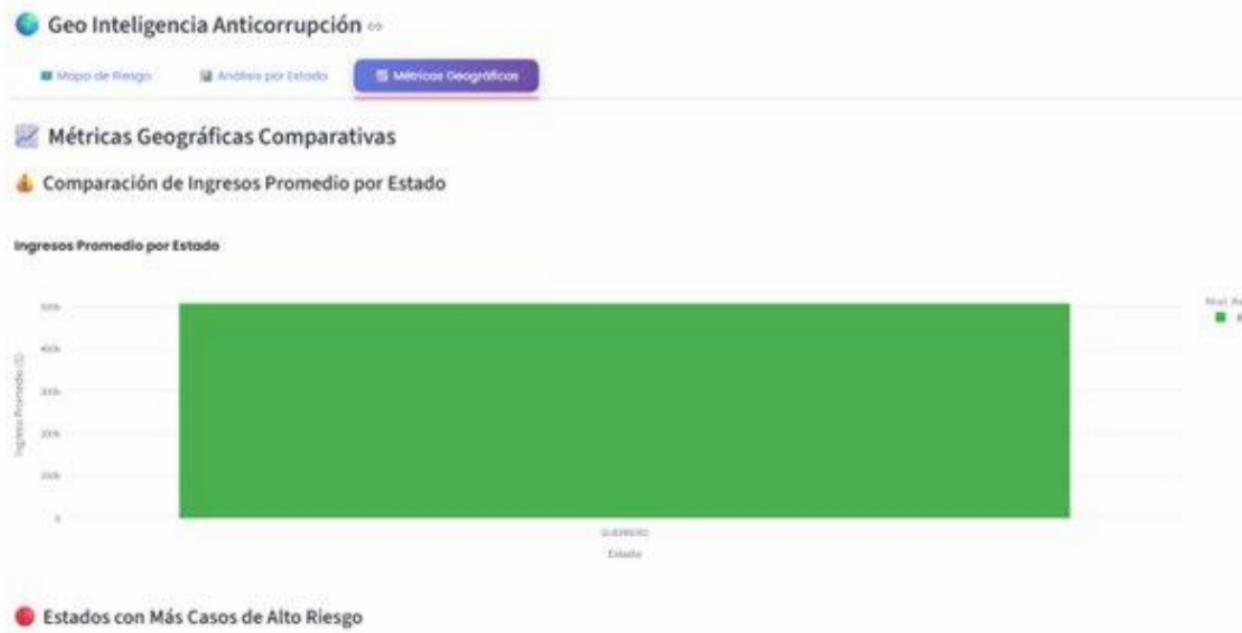


Figure 5.14: Comparativa de ingresos promedio por estado

5.6 Evolución Patrimonial (CORE del Proyecto)

La pestaña de “Evolución Patrimonial” es el núcleo central de AURORA ETHICS. Esta sección tiene 4 pestañas internas que permiten un análisis exhaustivo.

5.6.1 Tabla Principal

La primera pestaña presenta la tabla principal. La visualización inicial muestra por defecto los servidores públicos con **RIESGO ALTO**:

The screenshot shows the 'Evolución Patrimonial y Análisis de Riesgos' dashboard with the 'Tabla Principal' tab selected. At the top, there are four tabs: 'Tabla Principal' (selected), 'Buscador Inteligente', 'Visualizaciones', and 'Alertas de Corrupción'. Below the tabs, a summary box displays 'Total de servidores públicos analizados: 99'. The main area has three filters: 'Estado' (set to 'TODOS'), 'Nivel de Riesgo' (set to 'ALTO'), and '¿Tiene valores cero?' (set to 'TODOS'). The table header includes columns: 'NombreCompleto', 'Estado', 'Dependencia', 'Cargo', 'n_declaraciones', 'Decla_sin0SP', 'Decla_sin0OI', 'Decla_sin0TI', and 'ingreso_n'. The table body contains 99 rows of data, each representing a public servant with their name, state, dependency, position, and various metrics. A button at the bottom left allows 'Descargar Tabla Completa (Excel)'.

Figure 5.15: Tabla de evolución patrimonial - Filtro de Riesgo Alto

Para visualizar la tabla completa con todos los niveles de riesgo, hay que quitar el filtro de “ALTO”:

The screenshot shows the same dashboard as Figure 5.15, but with the 'Nivel de Riesgo' filter removed, resulting in 99 entries instead of 9. The table structure is identical to Figure 5.15, showing 99 rows of data for public servants across various states and dependencies with their respective metrics.

Figure 5.16: Tabla principal completa con todas las métricas calculadas

Métricas y Cálculos Incluidos en la Tabla Principal:

La tabla incluye las siguientes columnas calculadas automáticamente:

1. **NombreCompleto**: Identificación del servidor público
2. **n_declaraciones**: Número total de declaraciones presentadas
3. **Decla_sin0SP**: Declaraciones con ingreso del servidor público mayor a cero
4. **Decla_sin0OI**: Declaraciones con otros ingresos distintos de cero

5. **Decla_sin0TI:** Declaraciones con ingresos totales no nulos
6. **ingreso_sp_promedio:** Promedio de ingresos como servidor público (excluyendo ceros)
7. **max_ingresos_SP:** Máximo ingreso del servidor público registrado
8. **min_ingresos_SP:** Mínimo ingreso del servidor público registrado
9. **otros_ingresos_promedio:** Promedio de otros ingresos (excluyendo ceros)
10. **max_ingresos_OI:** Máximo de otros ingresos registrado
11. **min_ingresos_OI:** Mínimo de otros ingresos registrado
12. **ingresos_totales_promedio:** Promedio de todos los ingresos (excluyendo ceros)
13. **max_ingresos_totales:** Máximo ingreso total registrado
14. **min_ingresos_totales:** Mínimo ingreso total registrado
15. **tiene_valores_cero:** Indicador binario de existencia de valores en cero
16. **No_bienesInm:** Número de bienes inmuebles (sin duplicados)
17. **Importe_bienesInm:** Valor total de bienes inmuebles
18. **No_autos:** Número de vehículos (sin duplicados)
19. **Importe_autos:** Valor total de vehículos
20. **No_inversiones:** Número de inversiones (sin duplicados)
21. **Importe_inversiones:** Valor total de inversiones
22. **No_deudas:** Número de adeudos (sin duplicados)
23. **Importe_adeudos_total:** Valor total de adeudos
24. **delta_max_abs:** Máximo incremento absoluto en ingresos
25. **delta_max_pct:** Máximo incremento porcentual en ingresos
26. **alerta_incremento_inusual:** Alerta por incremento 100%
27. **fecha_ultima_actualizacion:** Fecha de la última declaración
28. **Estado:** Estado de adscripción
29. **Dependencia:** Institución donde labora
30. **Cargo:** Puesto del servidor público
31. **AnioDeclaracion:** Año de la declaración
32. **pct_incremento_anual:** Porcentaje de incremento anual

33. **alertas_detalle:** Detalle textual de todas las alertas
34. **nivel_riesgo_corrupcion:** Clasificación ALTO/MEDIO/BAJO
35. **alerta_ingresos:** Alerta por incremento inusual de ingresos
36. **alerta_otros_ingresos:** Alerta por otros ingresos desproporcionados
37. **alerta_adeudos:** Alerta por adeudos elevados
38. **alerta_bienes:** Alerta por bienes inmuebles desproporcionados
39. **alerta_global_riesgo:** Alerta global que consolida todas las anteriores

Debajo de la tabla principal se encuentra un botón para descargar la tabla completa calculada en formato Excel, preservando todas las métricas y cálculos.

The screenshot shows a dashboard titled "Evolución Patrimonial y Análisis de Riesgos". At the top, there are four tabs: "Tabla Principal" (highlighted in blue), "Buscador Inteligente", "Visualizaciones", and "Alertas de Corrupción". Below the tabs, a section titled "Tabla Principal de Evolución Patrimonial" displays a message: "Total de servidores públicos analizados: 99". There are three dropdown filters: "Estado" (set to "Todos"), "Nivel de Riesgo" (set to "Choose options"), and "¿Tiene valores cero?" (set to "Todos"). A message "Registros mostrados: 99" is shown above a table. The table has columns: NombreCompleto, Estado, Dependencia, Cargo, n_declaraciones, Decla_sindSP, Decla_sindO, Decla_sindIT, and ingresos. The data consists of 11 rows of public servant information from Guerrero.

NombreCompleto	Estado	Dependencia	Cargo	n_declaraciones	Decla_sindSP	Decla_sindO	Decla_sindIT	ingresos
ABBY YOUTH CASTAÑEDA ADAME	GUERRERO	INSTITUTO DE TRANSPARENCIA, ACCESO A LA INFORMACIÓN Y PROTECCIÓN DE DATOS. DIRECTORA	DIRECTORA	1	1	0	1	
ALFREDO MORALES HIRIENDA	GUERRERO	TRIBUNAL DE JUSTICIA ADMINISTRATIVA DEL ESTADO DE GUERRERO	SECRETARIO DE ACUERDOS	1	1	0	1	
ANAHARRA ABIGAIL FUENTES LOPEZ	GUERRERO	INSTITUTO DE TRANSPARENCIA, ACCESO A LA INFORMACIÓN Y PROTECCIÓN DE DATOS. CONTADORES	CONTADOR	1	1	0	1	
ANIA CRISTAL CASAMONICA PETATAN	GUERRERO	AYUNTAMIENTO DE JUCHITÁN GUERRERO	SINDICA	1	1	0	1	
ANA GABRIEL ANTUNEZ NUÑEZ	GUERRERO	INSTITUTO DE TRANSPARENCIA, ACCESO A LA INFORMACIÓN Y PROTECCIÓN DE DATOS. SECRETARIA DE ACUERDOS	SECRETARIA DE ACUERDOS	1	1	0	1	
ANA MELI AGUILAR ZURITA	GUERRERO	AYUNTAMIENTO DE JUCHITÁN GUERRERO	REGIDORA	1	1	0	1	
ANABEL SOLÍS RIOS	GUERRERO	INSTITUTO DE TRANSPARENCIA, ACCESO A LA INFORMACIÓN Y PROTECCIÓN DE DATOS. ENCARGADA DE LA PLATAFOR	ENCARGADA DE LA PLATAFOR	1	1	0	1	
ARACELI GONZALEZ MARCHAN	GUERRERO	TRIBUNAL DE JUSTICIA ADMINISTRATIVA DEL ESTADO DE GUERRERO	ASESOR CIUDADANO	1	1	0	1	
ARNOLD GARCIA PEREZ	GUERRERO	INSTITUTO DE TRANSPARENCIA, ACCESO A LA INFORMACIÓN Y PROTECCIÓN DE DATOS. VERIFICADOR	VERIFICADOR	1	1	0	1	
ARTURO BLANCO SANCHEZ	GUERRERO	TRIBUNAL DE JUSTICIA ADMINISTRATIVA DEL ESTADO DE GUERRERO	ASESOR CIUDADANO	1	1	0	1	
BENJAMIN MOJICA VALENCIA	GUERRERO	TRIBUNAL DE JUSTICIA ADMINISTRATIVA DEL ESTADO DE GUERRERO	CONFIANZA	1	1	0	1	

Figure 5.17: Tabla de evolución patrimonial con opciones de descarga

5.6.2 Buscador Inteligente

La segunda pestaña contiene el **Buscador Inteligente de Servidores Públicos**:



Figure 5.18: Buscador inteligente - Interfaz de búsqueda

Este buscador permite:

- Búsqueda por nombre completo
- Búsqueda por cargo o dependencia
- Filtrado por nivel de riesgo
- Filtrado por estado
- Búsqueda combinada con múltiples criterios

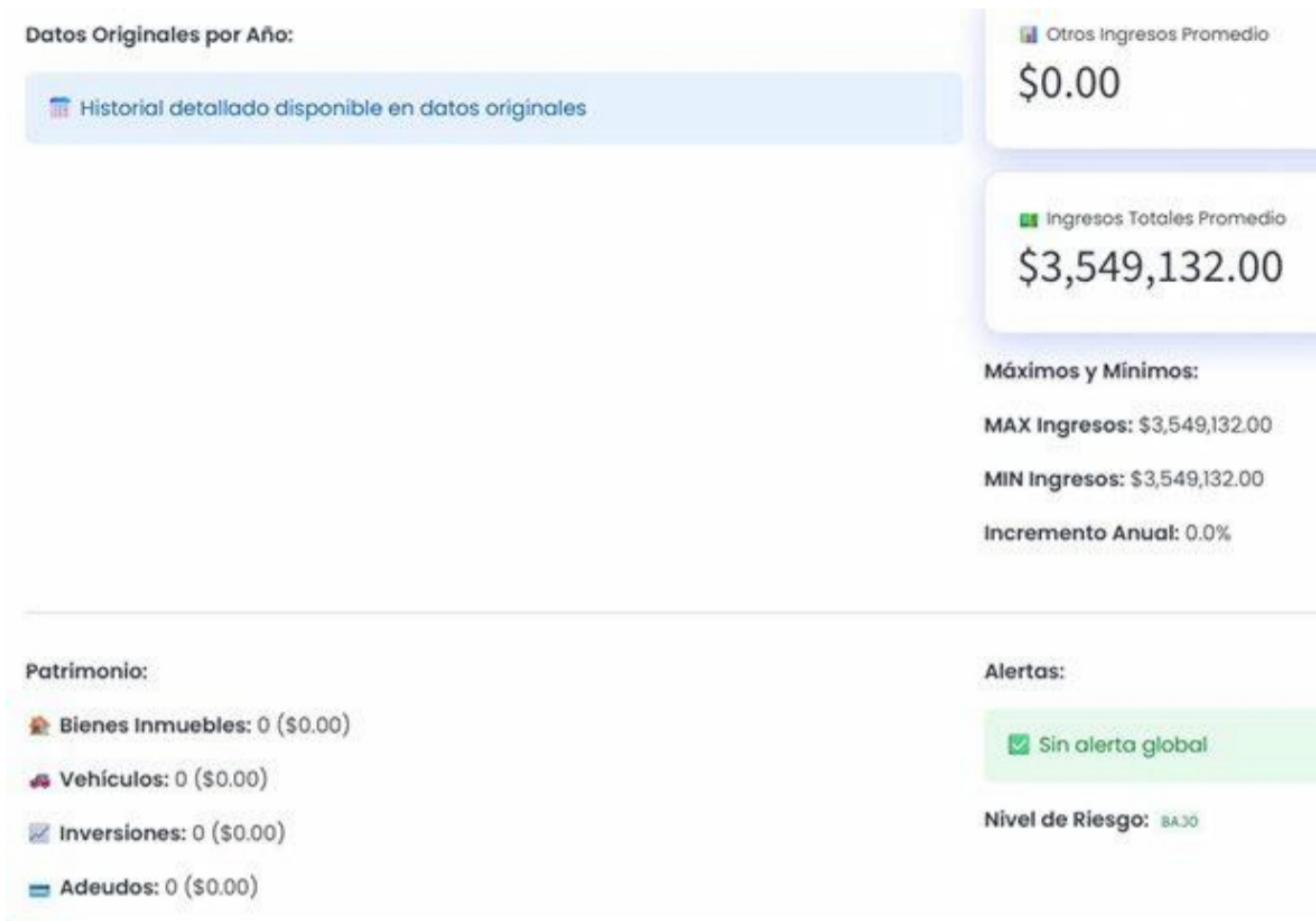


Figure 5.19: Resultados del buscador inteligente

5.6.3 Top 20 Casos de Alto Riesgo

La tercera pestaña presenta el **Top 20 de Casos de Alto Riesgo**:

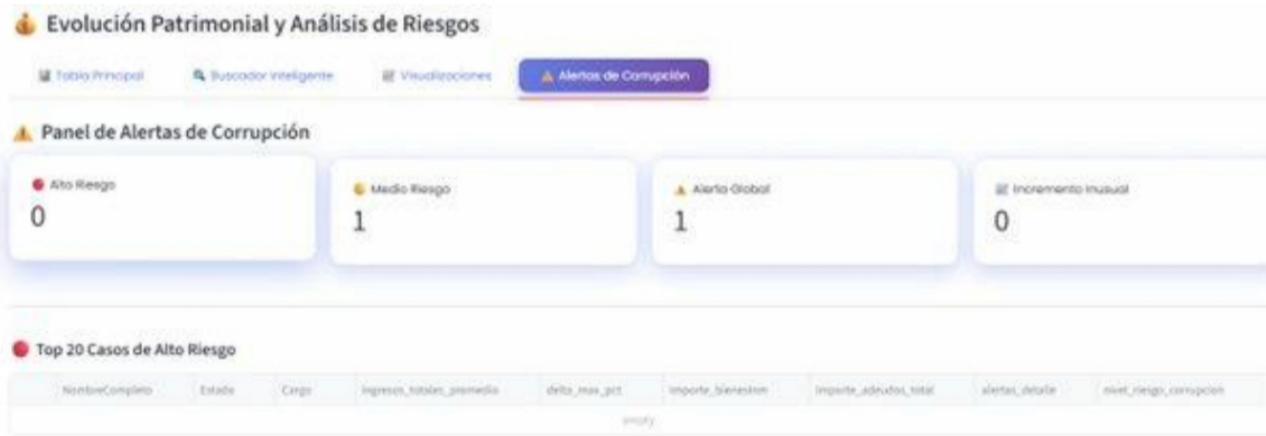


Figure 5.20: Top 20 casos de alto riesgo

Esta vista prioriza los casos más críticos que requieren investigación inmediata, ordenados por:

- Puntuación de riesgo
- Magnitud de incremento patrimonial
- Número y severidad de alertas
- Desproporción entre ingresos y patrimonio

5.6.4 Reporte de Métricas

La cuarta pestaña contiene el **Reporte de Métricas**:

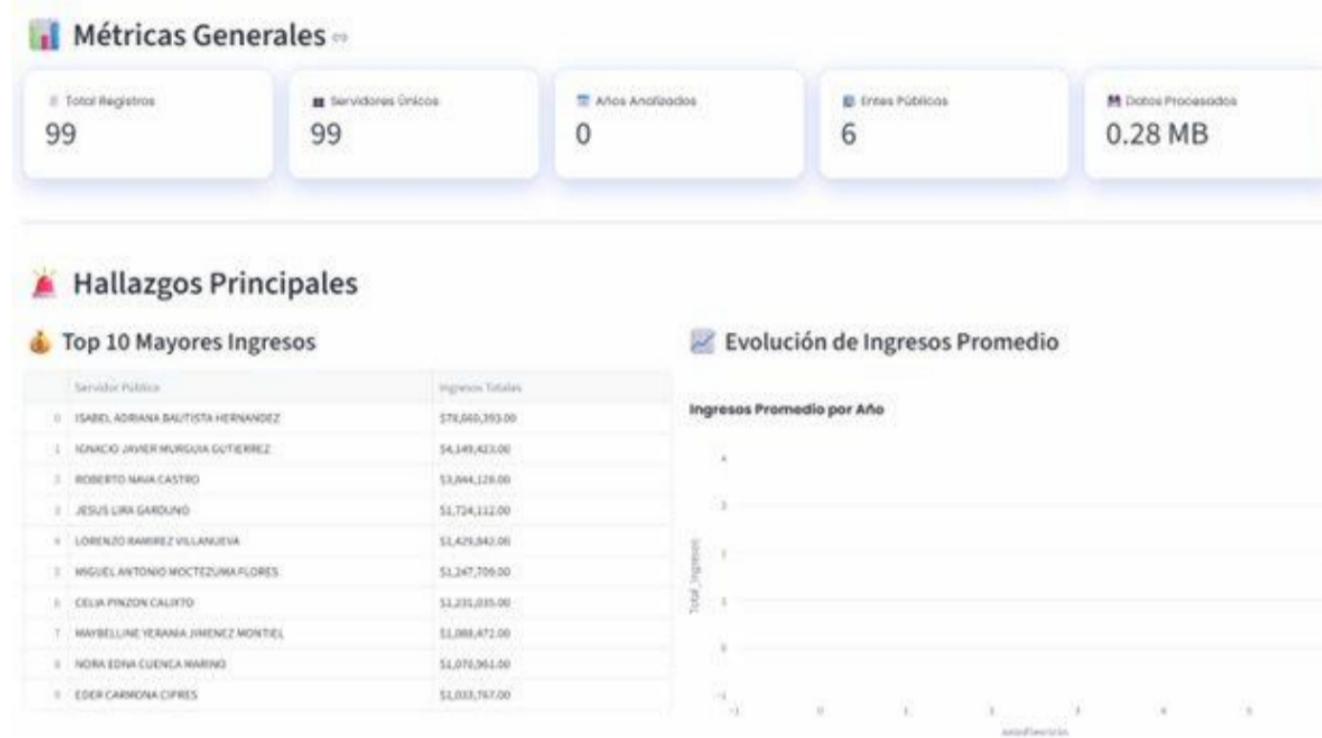


Figure 5.21: Reporte de métricas consolidadas

Este reporte incluye:

- Estadísticas generales del análisis
- Distribución de niveles de riesgo
- Promedios y medianas de indicadores clave
- Comparativas entre estados
- Tendencias temporales

5.7 Nepotismo

La sección de “Nepotismo” implementa algoritmos de detección de redes y relaciones familiares sospechosas.

5.7.1 Red de Relaciones



Figure 5.22: Red de relaciones - Detección de nepotismo

El sistema utiliza:

- Algoritmos de similitud de nombres (Levenshtein distance)
- Análisis de grafos para identificar clusters familiares
- Detección de coincidencias en apellidos y dependencias

5.7.2 Estadísticas del Grafo

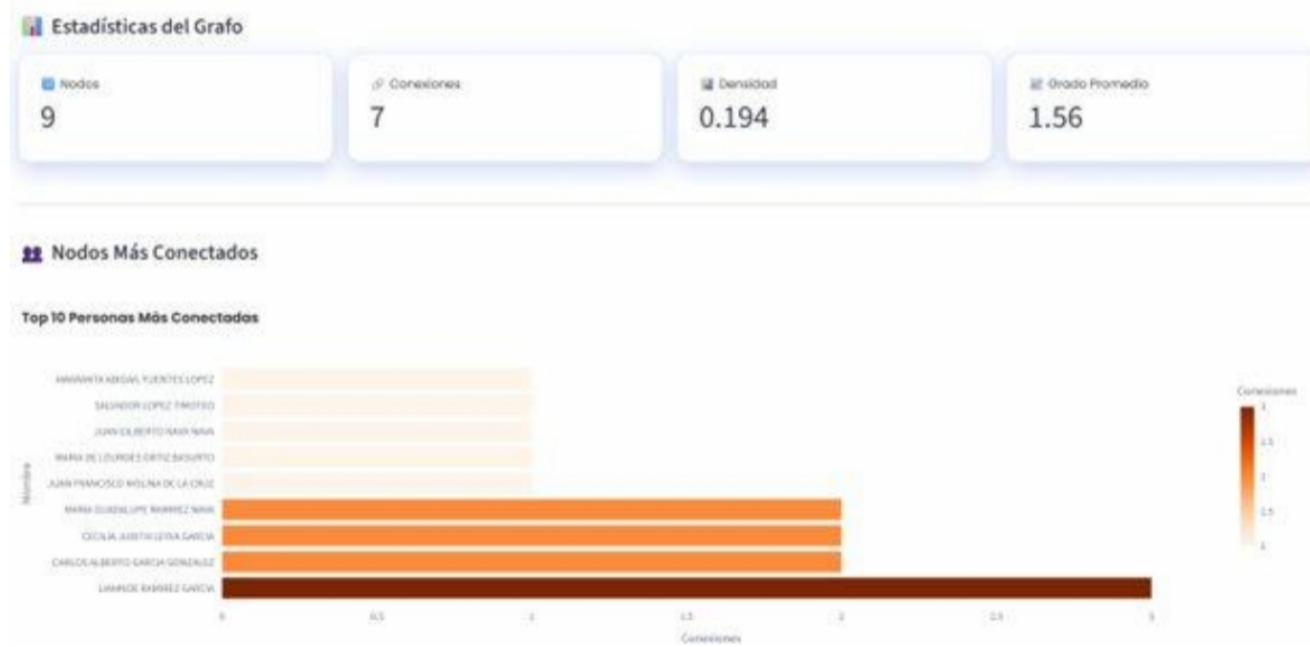


Figure 5.23: Estadísticas del grafo de relaciones

Métricas incluidas:

- Número de nodos (servidores públicos)
- Número de aristas (relaciones detectadas)
- Componentes conectados

- Densidad de la red
- Grado promedio de conexiones

5.7.3 Distribución por Tipo de Coincidencia

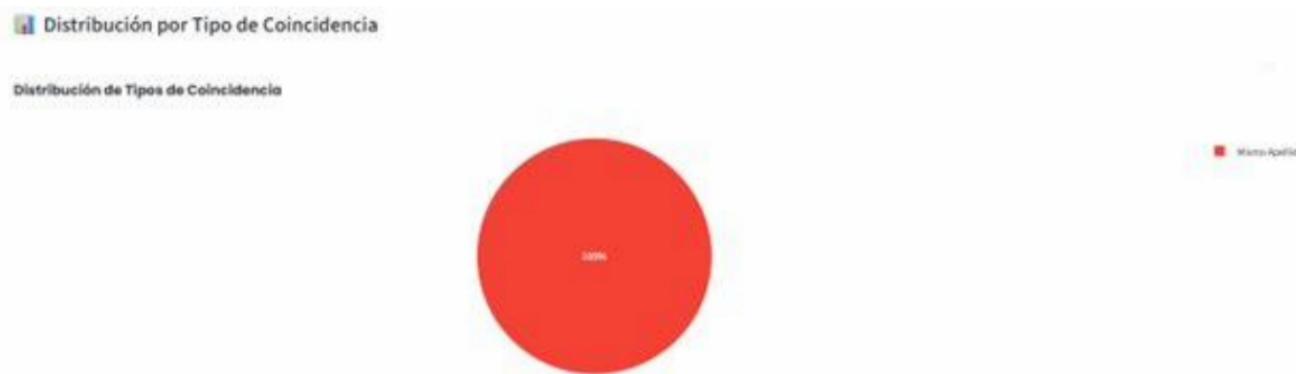


Figure 5.24: Distribución por tipo de coincidencia en nepotismo

5.7.4 Top 20 Casos de Similitud

Top 20 Coincidencias por Similitud				
Número 1	Número 2	Similitud (%)	Ente Público	Tipo
29 ARTURO BLANCO SANCHEZ	BERTINA GAMA SANCHEZ	60	TRIBUNAL DE JUSTICIA ADMINISTRATIVA DEL ESTADO DE GUERRERO	Mismo Apellido
18 PETRA PEREZ HERRERA	GUADALUPE PEREZ HERNANDEZ	50	AYUNTAMIENTO DE JUCHITAN GUERRERO	Mismo Apellido
19 PAOLA GARCIA IROS	SABAD GARCIA PRUDENTE	50	AYUNTAMIENTO DE JUCHITAN GUERRERO	Mismo Apellido
15 ARNOLD GARCIA PEREZ	JAIMIE GARCIA MORENO	50	INSTITUTO DE TRANSPARENCIA, ACCESO A LA INFORMACION Y PROTECCION DE DATOS PERSONALES DEL ESTADO DE GUERRERO	Mismo Apellido
2 MARIA GUADALUPE RAMIREZ NAVAR	LIAHNE RAMIREZ GARCIA	50	INSTITUTO DE TRANSPARENCIA, ACCESO A LA INFORMACION Y PROTECCION DE DATOS PERSONALES DEL ESTADO DE GUERRERO	Mismo Apellido
12 KEVIN GUADALUPE ORTEGA JIMENEZ	MA. GUADALUPE FRANCO CORONEL	50	INSTITUTO DE TRANSPARENCIA, ACCESO A LA INFORMACION Y PROTECCION DE DATOS PERSONALES DEL ESTADO DE GUERRERO	Mismo Apellido
22 EUNICE RAMIREZ BARAJAS	LORENZO RAMIREZ VILLANUEVA	50	TRIBUNAL DE JUSTICIA ADMINISTRATIVA DEL ESTADO DE GUERRERO	Mismo Apellido
12 KEVIN GUADALUPE ORTEGA JIMENEZ	JORGE IVAN NARANJO JIMENEZ	50	INSTITUTO DE TRANSPARENCIA, ACCESO A LA INFORMACION Y PROTECCION DE DATOS PERSONALES DEL ESTADO DE GUERRERO	Mismo Apellido
6 LIAMINDE RAMIREZ GARCIA	CECILIA JUDITH LEYVA GARCIA	50	INSTITUTO DE TRANSPARENCIA, ACCESO A LA INFORMACION Y PROTECCION DE DATOS PERSONALES DEL ESTADO DE GUERRERO	Mismo Apellido
8 VERENICE TAPIA ROMAN	LILIANA UZETH VENTURA ROMAN	50	INSTITUTO DE TRANSPARENCIA, ACCESO A LA INFORMACION Y PROTECCION DE DATOS PERSONALES DEL ESTADO DE GUERRERO	Mismo Apellido

Figure 5.25: Top 20 casos de coincidencia por similitud

5.7.5 Entes con Mayor Coincidencia

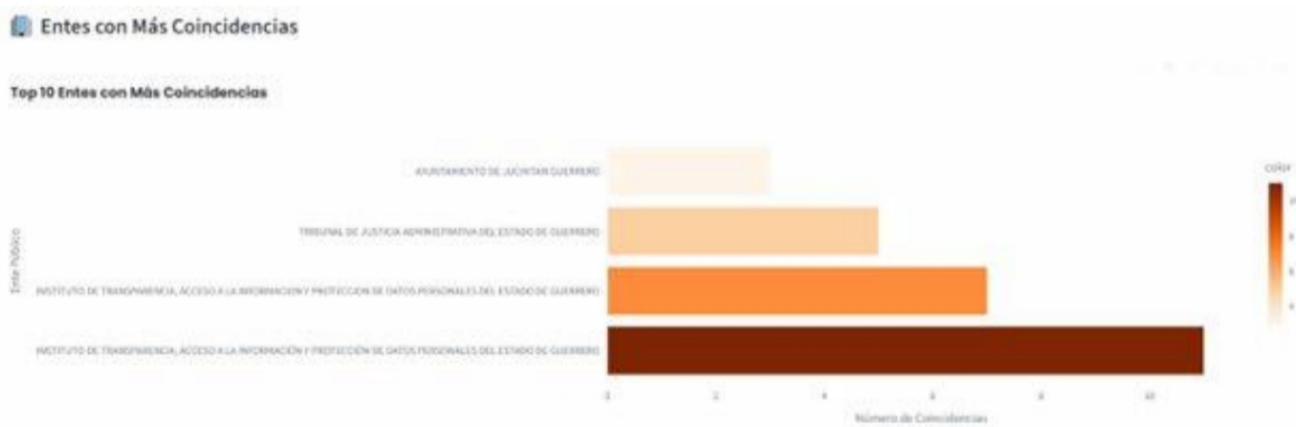


Figure 5.26: Entes públicos con más coincidencias de nepotismo

5.7.6 Tabla de Casos Sospechosos

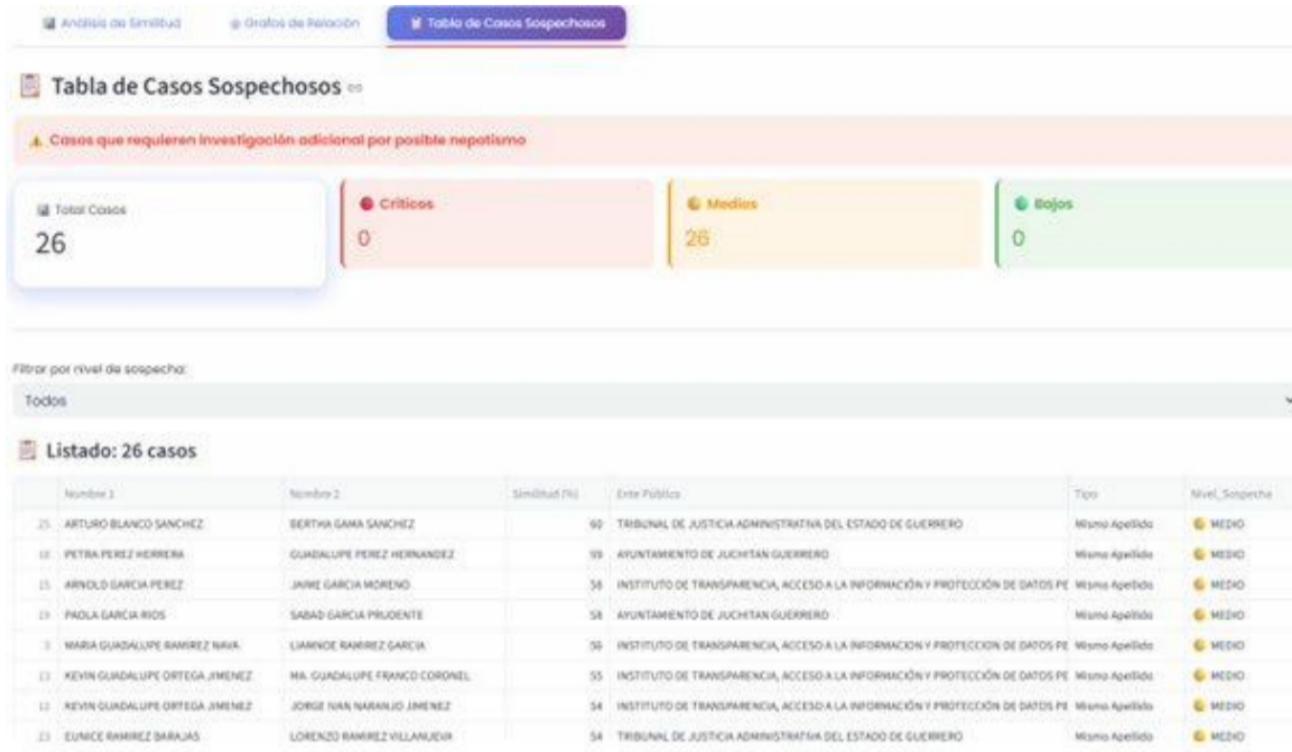


Figure 5.27: Tabla de casos sospechosos de nepotismo

5.8 Sistema de Denuncias Ciudadanas

La sección de “**Denuncias**” es un piloto sencillo adicional que incorpora un sistema de denuncias ciudadanas.

5.8.1 Registrar Nueva Denuncia

Figure 5.28: Formulario para registrar nueva denuncia

El sistema permite:

- Registro anónimo o identificado
- Categorización de la denuncia
- Adjuntar evidencia textual
- Selección de entidad y dependencia
- Descripción detallada del caso

5.8.2 Historial de Denuncias

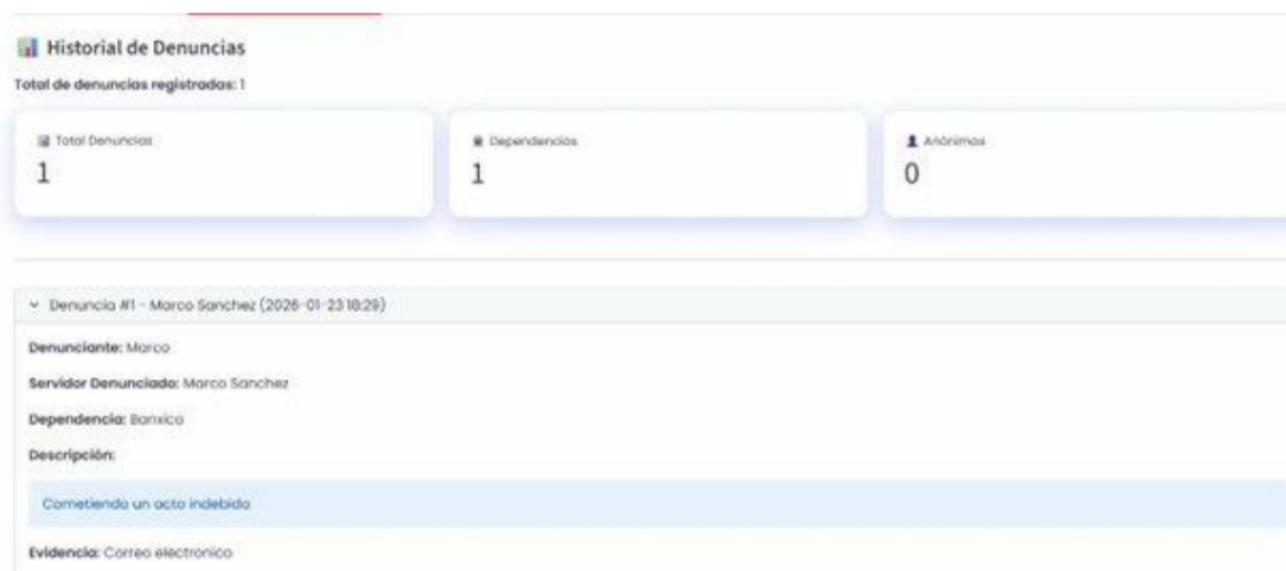


Figure 5.29: Historial de denuncias registradas

5.8.3 WordCloud de Denuncias

Para analizar las denuncias de forma visual, se genera una nube de palabras:



Figure 5.30: WordCloud de términos en denuncias

Esta visualización permite identificar rápidamente:

- Términos más frecuentes en las denuncias
- Patrones temáticos recurrentes
- Áreas de mayor preocupación ciudadana

5.9 Reporte Ejecutivo

Finalmente, la sección de “Reporte Ejecutivo” consolida todos los hallazgos:



Figure 5.31: Pantalla de reporte ejecutivo

El reporte ejecutivo muestra los hallazgos principales del análisis:

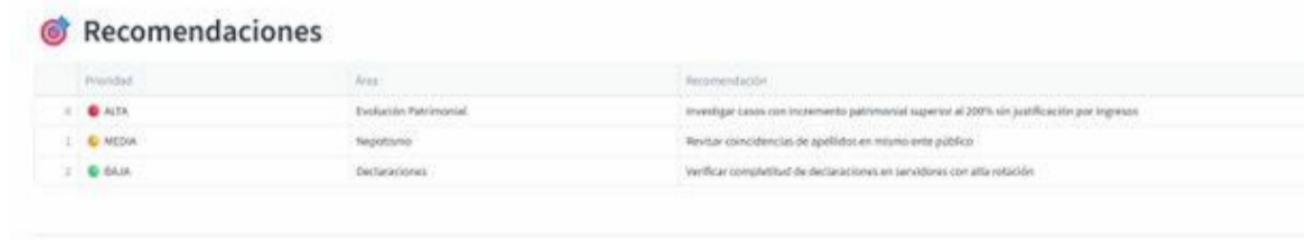


Figure 5.32: Recomendaciones del reporte ejecutivo

Este reporte incluye:

- Resumen ejecutivo de hallazgos
- Casos críticos identificados
- Recomendaciones de acción
- Estadísticas consolidadas
- Priorización de investigaciones
- Métricas de riesgo agregadas

Chapter 6

Escalabilidad: Implementación con MongoDB

6.1 Ejecución Local vs. Producción

Esta ejecución funciona de forma **local** y siguiendo las instrucciones proporcionadas se puede ejecutar el programa sin necesidad de infraestructura adicional. Sin embargo, para implementar y expandir su uso en un entorno de producción, especialmente cuando se manejan volúmenes masivos de datos y se requiere acceso concurrente de múltiples usuarios, se recomienda utilizar bases de datos especializadas.

6.2 ¿Por Qué MongoDB?

MongoDB es una base de datos NoSQL orientada a documentos que ofrece:

- **Escalabilidad horizontal:** Capacidad de distribuir datos entre múltiples servidores
- **Flexibilidad de esquema:** Ideal para datos JSON semi-estructurados como los de la PDN
- **Alto rendimiento:** Consultas rápidas incluso con millones de registros
- **Replicación y alta disponibilidad:** Garantiza que el sistema esté siempre operativo
- **Consultas avanzadas:** Permite agregaciones complejas y análisis en tiempo real

6.3 Arquitectura Propuesta

Arquitectura Sugerida

1. **Capa de Presentación:** Aplicación web Streamlit (existente)
2. **Capa de Aplicación:** Backend Python con lógica de negocio
3. **Capa de Datos:** MongoDB para almacenamiento persistente
4. **Capa de Caché:** Redis para consultas frecuentes (opcional)
5. **Capa de Seguridad:** Autenticación y autorización de usuarios

6.4 Instalación de MongoDB

6.4.1 Paso 1: Descargar MongoDB Community Edition

Acceder a la página oficial: <https://www.mongodb.com/try/download/community>

Seleccionar la versión apropiada para su sistema operativo (Windows, macOS, Linux).

6.4.2 Paso 2: Instalar MongoDB

Para Windows:

1. Ejecutar el instalador descargado
2. Seleccionar “Complete” en el tipo de instalación
3. Configurar MongoDB como servicio de Windows (recomendado)
4. Completar la instalación

Para macOS (usando Homebrew):

```
brew tap mongodb/brew
brew install mongodb-community@7.0
```

Para Linux (Ubuntu/Debian):

```
wget -qO - https://www.mongodb.org/static/pgp/server-7.0.asc | sudo
apt-key add -
echo "deb [ arch=amd64,arm64 ] https://repo.mongodb.org/apt/ubuntu
jammy/mongodb-org/7.0 multiverse" | sudo tee /etc/apt/sources.list.d/mongodb-org-7.0.list
sudo apt-get update
sudo apt-get install -y mongodb-org
```

6.4.3 Paso 3: Verificar la Instalación

Abrir una terminal o CMD y ejecutar:

```
mongo --version
```

Debería mostrar la versión de MongoDB instalada.

6.4.4 Paso 4: Iniciar el Servicio MongoDB

Windows:

```
net start MongoDB
```

macOS:

```
brew services start mongodb-community@7.0
```

Linux:

```
sudo systemctl start mongod
sudo systemctl enable mongod
```

6.5 Integración con AURORA ETHICS

6.5.1 Paso 5: Instalar PyMongo

Instalar la librería de Python para MongoDB:

```
pip install pymongo dnspython
```

6.5.2 Paso 6: Configurar la Conexión

En el código de AURORA_ETHICS.py, agregar la configuración de conexión a MongoDB:

```
import pymongo
from pymongo import MongoClient

# Conectar a MongoDB local
client = MongoClient('mongodb://localhost:27017/')

# Crear/seleccionar base de datos
db = client['aurora_ethics_db']

# Crear/seleccionar colecciones
declaraciones = db['declaraciones']
metricas = db['metricas_calculadas']
alertas = db['alertas_riesgo']
denuncias = db['denuncias_ciudadanas']
```

6.5.3 Paso 7: Insertar Datos en MongoDB

Modificar la función de procesamiento de JSON para guardar en MongoDB:

```
def guardar_en_mongodb(df_consolidado):
    """Guarda el DataFrame en MongoDB"""
    registros = df_consolidado.to_dict('records')

    # Insertar en MongoDB
    declaraciones.insert_many(registros)

    # Crear índices para búsquedas rápidas
    declaraciones.create_index("NombreCompleto")
    declaraciones.create_index("Estado")
    declaraciones.create_index("nivel_riesgo_corrupcion")

    return True
```

6.5.4 Paso 8: Consultar Datos desde MongoDB

Modificar las funciones de consulta para leer desde MongoDB:

```
def obtener_alto_riesgo():
    """Obtiene casos de alto riesgo desde MongoDB"""
    resultados = declaraciones.find(
        {"nivel_riesgo_corrupcion": "ALTO"}
    ).sort("delta_max_pct", -1)

    return pd.DataFrame(list(resultados))
```

6.6 Ventajas de la Implementación con MongoDB

- Persistencia de Datos:** Los datos quedan almacenados permanentemente, no solo en memoria
- Consultas Rápidas:** Índices optimizados permiten búsquedas instantáneas incluso con millones de registros
- Escalabilidad:** Capacidad de manejar todo el dataset nacional sin limitaciones de memoria RAM
- Análisis Histórico:** Posibilidad de comparar declaraciones entre diferentes años
- API REST:** Facilita crear APIs para que otros sistemas consuman los datos
- Respaldos Automáticos:** Configuración de backups programados
- Acceso Concurrente:** Múltiples usuarios pueden consultar simultáneamente
- Agregaciones Complejas:** Pipeline de agregación para análisis avanzados

6.7 Seguridad y Autenticación

Para un entorno de producción, es crucial implementar:

- **Autenticación de usuarios:** MongoDB con usuario y contraseña
- **Conexión encriptada:** Usar TLS/SSL
- **Control de acceso basado en roles (RBAC):** Diferentes permisos por tipo de usuario
- **Auditoría:** Registro de todas las consultas y modificaciones
- **Firewall:** Limitar acceso a IP específicas

Ejemplo de Conexión Segura

```
client = MongoClient(  
    'mongodb://usuario:contraseña@servidor:27017/' ,  
    tls=True,  
    tlsCAFile='/ruta/certificado.pem'  
)
```

6.8 MongoDB Atlas (Nube)

Para una implementación en la nube sin gestionar infraestructura:

1. Crear cuenta en MongoDB Atlas: <https://www.mongodb.com/cloud/atlas/register>
2. Crear un cluster gratuito o de pago según necesidades
3. Obtener la cadena de conexión
4. Configurar lista blanca de IPs
5. Conectar desde AURORA ETHICS usando la URI proporcionada

Ventajas de Atlas:

- Sin instalación ni mantenimiento de servidores
- Respaldos automáticos
- Escalado automático
- Monitoreo en tiempo real
- Alta disponibilidad global

Agradecimientos

Queremos expresar nuestro más profundo agradecimiento a la **Secretaría Ejecutiva del Sistema Nacional Anticorrupción (SESNÁ)** por su invaluable apoyo, orientación y confianza en este proyecto.

De manera especial, extendemos nuestro reconocimiento a las siguientes personas, cuya guía, disposición y experiencia fueron fundamentales para el desarrollo de AURORA ETHICS:

- **Dra. Vania Pérez Morales**
- **Ángel López Gálvan**
- **Erik Vinicio Montero Hernández**
- **Benito Perdomo Vergara**
- **Andrés López Ramón**
- **Mauro Talamantes Villagrana**
- **Marissa Martínez García**
- **José Ángel Fernández Hernández**
- **Sergio Alberto Rodríguez Segura**

Su compromiso con la transparencia, la ética en el servicio público y la lucha contra la corrupción ha sido una inspiración constante para nosotros. Gracias por compartir su conocimiento, por sus valiosas retroalimentaciones y por creer en el potencial de la tecnología como herramienta de cambio social.

Este proyecto no habría sido posible sin su apoyo. Esperamos que AURORA ETHICS sea una contribución significativa a los esfuerzos nacionales por fortalecer la rendición de cuentas y construir un México más transparente y justo.

“La transparencia es el mejor antídoto contra la corrupción.”

Con gratitud y compromiso,
Marco Antonio Sánchez Cedillo (M.A.S.C.)
Saúl Ávila Hernández (S.A.H.)