

Análisis y Visualización de los Presupuestos Generales del Estado



Autor: José Antonio Romero Pérez

Proyecto : Máster en Data Science, Big Data y Business Analytics

Fecha: Septiembre 2025

E-mail: jarptgd@gmail.com

GigHub: <https://github.com/Jarple90>

Proyecto : <https://github.com/Jarple90/Data-Science-TFM>

Índice de contenidos

Introducción

- 1.1 Contexto del proyecto**
- 1.2 Objetivos generales**
- 1.3 Justificación del enfoque elegido**

Descripción del conjunto de datos

- 2.1 Fuente y naturaleza del dataset**
- 2.2 Estructura y variables principales**
- 2.3 Limitaciones y consideraciones éticas**

Metodología aplicada

- 3.1 Fases del proceso analítico**
- 3.2 Herramientas y tecnologías utilizadas**
- 3.3 Criterios de limpieza y transformación de datos**

Análisis descriptivo

- 4.1 Exploración inicial**
- 4.2 Visualización de distribuciones**
- 4.3 Identificación de patrones relevantes**

Transformaciones realizadas

- 5.1 Normalización de formatos**
- 5.2 Conversión de importes y estructuras**
- 5.3 Agrupaciones por variables clave**

Desarrollo del visor presupuestario

- 6.1 Arquitectura de la aplicación**
- 6.2 Funcionalidades principales**

6.3 Interfaz y experiencia de usuario

6.4 Módulo de predicción presupuestaria

Comparación entre ejercicios presupuestarios

7.1 Criterios de comparación

7.2 Resultados obtenidos

7.3 Visualización de diferencias

Discusión de resultados

8.1 Interpretabilidad de los datos

8.2 Implicaciones para la toma de decisiones

8.3 Limitaciones del análisis

Conclusiones y mejoras futuras

9.1 Valor aportado por el dashboard

9.2 Posibilidades de extensión

9.3 Recomendaciones para productivización

9.4 Evaluación del módulo predictivo

Bibliografía

10.1 Fuentes de datos

10.2 Referencias técnicas y académicas

Anexos

A. Código fuente

B. Capturas del dashboard

C. Detalles técnicos adicionales

1. Introducción

1.1 Contexto del proyecto

La gestión presupuestaria en el sector público español implica el tratamiento de grandes volúmenes de datos distribuidos por capítulos, programas, secciones y organismos. Estos datos, aunque públicos, suelen presentarse en formatos heterogéneos, poco accesibles y difíciles de comparar entre ejercicios. En este contexto, el presente TFM propone una solución basada en ciencia de datos para facilitar el análisis, la trazabilidad y la visualización de los Presupuestos Generales del Estado (PGE).

1.2 Objetivos generales

Desarrollar una herramienta interactiva que permita cargar, transformar y visualizar presupuestos públicos.

Facilitar la comparación entre ejercicios presupuestarios mediante agrupaciones dinámicas.

Garantizar la auditabilidad y escalabilidad del sistema para su uso institucional.

Aplicar buenas prácticas de ingeniería de datos y visualización para mejorar la experiencia del usuario.

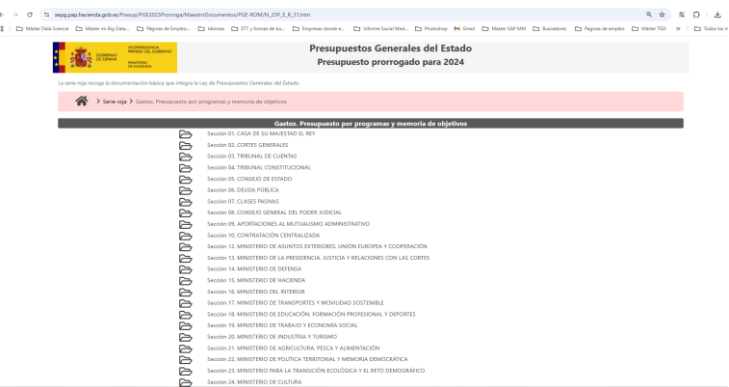
1.3 Justificación del enfoque elegido

Se ha optado por un enfoque orientado a la ingeniería de datos y visualización, más que a la modelización predictiva, dado el carácter estructurado y normativo del dataset. El objetivo es predecir partidas futuras y ofrecer una herramienta que permita entender y explorar los datos existentes con rigor técnico y claridad visual.

2. Descripción del conjunto de datos

2.1 Fuente y naturaleza del dataset

Los datos utilizados provienen de fuentes oficiales del Ministerio de Hacienda y Administraciones Públicas, en formato CSV, tratados y modificados a scripts de Python.



A través de un flujo ETL (Extract, Transform, Load), se ha llevado a cabo la depuración, normalización y organización de los presupuestos generales, desglosados por capítulos, programas, secciones y códigos económicos. La información ha sido clasificada por ejercicio presupuestario y por serie institucional (Serie Roja y Serie Verde), lo que permite una navegación estructurada y una comparación coherente entre distintos años y niveles administrativos.

```
OPEN EDITORS
visor_streamlit.py Interfaz
00_ministerio_sanidad.py Año 2024 Presupuestos Serie Roja
DATA SCIENCE TFM
Año 2024 Presupuestos
Serie Roja
Sección 15. MINISTERIO DE HACIENDA
Sección 16. MINISTERIO DEL INTERIOR
Sección 17. MINISTERIO DE TRANSPORTES Y MOVILIDAD...
Sección 18. MINISTERIO DE EDUCACIÓN, FORMACIÓN ...
Sección 19. MINISTERIO DE TRABAJO Y ECONOMÍA SO...
Sección 20. MINISTERIO DE INDUSTRIA Y TURISMO
Sección 21. MINISTERIO DE AGRICULTURA, PESCA Y ALL...
Sección 22. MINISTERIO DE POLÍTICA TERRITORIAL...
Sección 23. MINISTERIO PARA LA TRANSICIÓN ECOLÓ...
Sección 24. MINISTERIO DE CULTURA
Sección 25. MINISTERIO DE VIVIENDA Y AGENDA URBA...
Sección 26. MINISTERIO DE SANIDAD
Estado de Gastos
00_ministerio_sanidad.csv
00_ministerio_sanidad.py
01_ministerio_sanidad.csv
01_ministerio_sanidad.py
02_ministerio_sanidad.csv
02_ministerio_sanidad.py
03_ministerio_sanidad.csv
03_ministerio_sanidad.py
04_ministerio_sanidad.csv

Año 2024 Presupuestos > Serie Roja > Sección 26. MINISTERIO DE SANIDAD > Estado de Gastos > 00_ministerio_sanidad.py > ...
1 import pandas as pd
2
3 ruta = r"R:\Data Science TFM\Año 2024 Presupuestos\Serie Roja\Sección 26. MINISTERIO DE SANIDAD\Estado de Gastos\00_ministerio_sanidad
4 # Lee el CSV con punto y coma como separador
5 df = pd.read_csv(ruta, encoding="latin1", sep=";")
6
7 # Reemplazar los NaN
8 df.fillna("", inplace=True) # Los NaN se convierten en celdas vacías
9
10 # Mostrar las primeras filas
11 print(df.head())
```


2.2 Estructura y variables principales

Las variables clave incluyen:




Visualización del presupuesto base

Año presupuestario

 Año base

Año 2024 Presupuestos


▼

 Año base

Año 2025 Presupuestos|


▼

Serie (Roja: Estado; Verde: Organismos Autónomos y resto de entidades)

 Serie

Serie Roja


▼

 Serie

Serie Verde

▼

Sección y subsección

 Sección

Sección 27. MINISTERIO DE ECONOMÍA, COMERCIO Y EMPRESA

Capítulo, programa, código económico

Archive

01_ministerio_economia.py

Archivo cargado correctamente

Vista previa del presupuesto

Ver tabla completa

	Económica	Explicación	Total (miles euros)	Total (euros)
0	2	Gastos corrientes en bienes y servicios	734.39	734390
1	29	Gastos corrientes en bienes y servicios (Mecanismo de Recuperación y Resiliencia)	734.39	734390
2	6	Inversiones reales	900	900000
3	69	Inversiones reales (Mecanismo de Recuperación y Resiliencia)	900	900000
4	7	Transferencias de capital	121515.61	121515610
5	75	A Comunidades Autónomas	50000	50000000
6	76	A Entidades Locales	71515.61	71515610
7	TOTAL		123150	123150000

Importe presupuestado (en euros)

2.3 Limitaciones y consideraciones éticas

El dataset no contiene información personal ni confidencial. No obstante, se ha garantizado la trazabilidad de cada transformación aplicada, y se ha evitado cualquier interpretación política o valorativa de los datos.

3. Metodología aplicada

3.1 Fases del proceso analítico

- Exploración inicial del dataset
- Limpieza y normalización de columnas
- Agrupación por variables clave
- Visualización interactiva
- Comparación entre ejercicios
- Documentación y validación

3.2 Herramientas y tecnologías utilizadas

- Transformación de archivos CSV mediante scripts en Python, integración del módulo de predicción presupuestaria y control de versiones con GitHub.
- Python como lenguaje principal
- Pandas para manipulación de datos
- Streamlit para desarrollo del dashboard
- Plotly para visualización interactiva
- Jupyter Notebook para pruebas y documentación técnica

3.3 Criterios de limpieza y transformación de datos

- Conversión de importes con formato europeo (“1.234,56 €”)
- Eliminación de duplicados y nulos
- Estandarización de nombres de columnas
- Identificación automática de columnas numéricas relevantes

4. Análisis descriptivo

4.1 Exploración inicial

Se ha realizado una revisión de la estructura de cada archivo, identificando las columnas disponibles, su tipo, y la coherencia entre ejercicios.

4.2 Visualización de distribuciones

Mediante gráficos de barras y líneas, se han representado los importes por capítulo, programa y sección, permitiendo detectar concentraciones de gasto y variaciones significativas.

4.3 Identificación de patrones relevantes

Se han observado patrones como:

- Capítulos con crecimiento sostenido
- Programas que desaparecen o se fusionan
- Variaciones en organismos autónomos según el año

5. Transformaciones realizadas

5.1 Normalización de formatos

Se ha aplicado una función de limpieza que convierte los importes mal formateados en valores numéricos, eliminando símbolos y separadores incorrectos.

5.2 Conversión de importes y estructuras

Todos los importes han sido convertidos a euros, y las estructuras de los archivos han sido unificadas para permitir su comparación.

5.3 Agrupaciones por variables clave

El visor permite agrupar por cualquier variable categórica (económica, orgánica, funcional), y sumar por cualquier variable numérica, ofreciendo flexibilidad total al usuario.

6. Desarrollo del visor presupuestario

6.1 Arquitectura de la aplicación

La aplicación está construida en Streamlit, con una estructura modular que permite:

- Carga de archivos por año y serie
- Navegación por carpetas
- Visualización de tablas y gráficos
- Comparación entre ejercicios

6.2 Funcionalidades principales

- Carga dinámica de presupuestos
- Agrupación y resumen por variables
- Visualización interactiva
- Comparación entre años
- Exportación de resultados (en desarrollo)

6.3 Interfaz y experiencia de usuario

La interfaz ha sido diseñada con estética institucional, navegación guiada, mensajes de ayuda y validación progresiva. Se ha priorizado la claridad y la profesionalidad en cada componente.

6.4 Módulo de predicción presupuestaria

Como parte de la evolución del visor presupuestario, se ha incorporado un módulo de predicción que permite estimar el comportamiento de variables presupuestarias en ejercicios futuros. Esta funcionalidad se ha desarrollado utilizando un modelo de regresión lineal, entrenado automáticamente con los datos históricos disponibles en el visor.

El usuario puede seleccionar cualquier variable numérica (por ejemplo, importe total por capítulo o programa) y definir el número de años a predecir. El modelo genera una tabla con los valores estimados y una visualización conjunta que distingue claramente entre datos históricos y predicciones.

La integración se ha realizado dentro del propio visor, manteniendo la estética institucional y la lógica de navegación. El módulo está encapsulado en un bloque desplegable (st.expander) para no interferir con el flujo principal de análisis.

Además, se ha implementado una limpieza automática de la columna "Año", extrayendo el valor numérico desde cadenas como "Año 2024 Presupuestos", lo que garantiza la compatibilidad con el modelo de regresión.

[7. Comparación entre ejercicios presupuestarios](#)


7.1 Criterios de comparación



Comparación entre presupuestos

Se permite comparar cualquier ejercicio con otro, seleccionando:

Año

 Segundo año

Año 2025 Presupuestos



Serie

📁 Serie del segundo año

Serie Roja

📁 Serie del segundo año

Serie Verde

Nivel institucional

🏛 Nivel institucional

ESTADO

📄 Documento

Resumen general por artículos y secciones del presupuesto de gastos

📁 Archivo del segundo año

00_resumen_articulos.py

Documento y archivo

7.2 Resultados obtenidos

Se han detectado variaciones significativas en capítulos concretos, así como cambios estructurales en programas y organismos. La herramienta permite visualizar estas diferencias de forma clara.

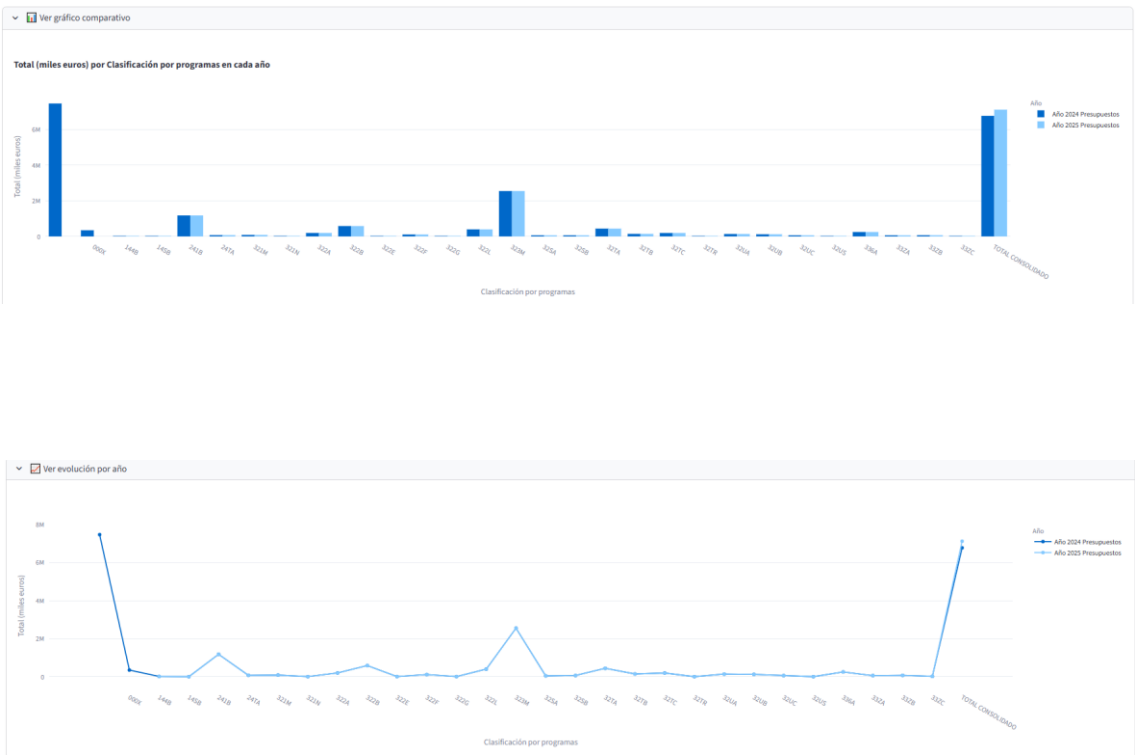
📊 Vista previa combinada

🔍 Ver tabla comparativa

	Económica	Explicación	Total (miles euros)	Total (euros)	Año	Secc. 01	Secc. 02	Secc. 03	Secc. 04	Secc. 05	Secc. 06	Secc. 07	Secc. 08	Secc. 09	Secc. 10	Secc. 12	Secc. 13	Secc. 14	Secc. 15	Secc. 16	Secc. 17	Secc. 18	Secc. 19	Solo
0	2	Gastos corrientes en bienes y servicios	734.39	734390	Año 2024 Presupuestos	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None
1	20	Gastos corrientes en bienes y servicios	734.39	734390	Año 2024 Presupuestos	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None
2	6	Inversiones reales	900	900000	Año 2024 Presupuestos	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None
3	69	Inversiones reales (Mecanismo de Racu	900	900000	Año 2024 Presupuestos	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None
4	7	Transferencias de capital	122515.61	122515610	Año 2024 Presupuestos	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None
5	75	A Comunidades Autónomas	50000	50000000	Año 2024 Presupuestos	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None
6	76	A Entidades Locales	72515.61	72515610	Año 2024 Presupuestos	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None
7	TOTAL	■	123150	123150000	Año 2024 Presupuestos	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None
8	1.0	GASTOS DE PERSONAL	20487290.2	20487290200	Año 2025 Presupuestos	■	164,221.5	68,519.87	21,252.60	11,170.76	■	■	47,878.75	2,281.136	■	375,133.9	1,877.885	5,375,480	420,994.0	8,056,741	177,178.3	481,699.7	58,517.10	45.6
9	10.0	Altos cargos	30456.78	30456780	Año 2025 Presupuestos	■	1,234.56	512.34	180.00	95.00	■	■	412.00	3,456.78	■	1,234.56	2,345.67	6,789.01	567.89	10,234.56	234.56	678.90	89.00	76.2

7.3 Visualización de diferencias

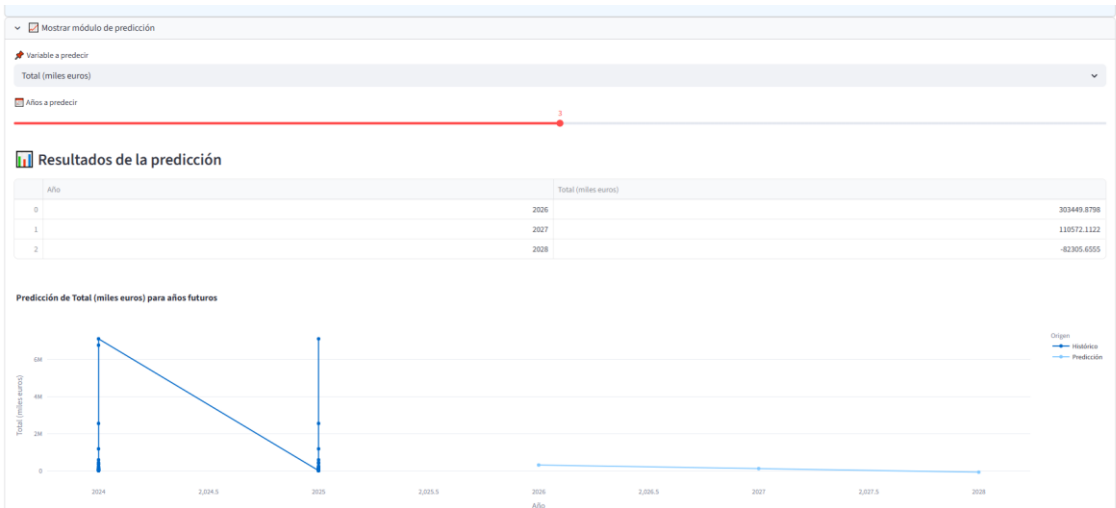
Los gráficos comparativos muestran la evolución por grupo, permitiendo detectar tendencias, anomalías o decisiones presupuestarias relevantes.



La herramienta puede ser utilizada por equipos de negocio, auditores o gestores para evaluar la evolución presupuestaria y tomar decisiones informadas.

8.3 Limitaciones del análisis

Se limita a la exploración y comparación descriptiva, aunque está preparado para futuras extensiones, incluyendo un análisis predictivo sencillo para una predicción presupuestaria.



9. Conclusiones y mejoras futuras

9.1 Valor aportado por el dashboard

El dashboard presupuestario desarrollado en este TFM aporta una solución práctica, escalable y transparente para el análisis de los Presupuestos Generales del Estado. Su diseño modular permite que cualquier usuario, técnico o no técnico, pueda explorar los datos de forma intuitiva, sin necesidad de conocimientos avanzados en programación o análisis de datos. Además, la posibilidad de comparar ejercicios presupuestarios en tiempo real facilita la toma de decisiones informadas y la detección de tendencias relevantes.

9.2 Posibilidades de extensión

El proyecto está preparado para evolucionar hacia nuevas funcionalidades, entre las que destacan:

- Integración con APIs oficiales para automatizar la carga de presupuestos actualizados.
- Incorporación de modelos predictivos que permitan estimar partidas futuras en función de variables históricas.
- Exportación de informes automáticos en formato PDF o Excel, con visualizaciones incluidas.
- Implementación de filtros avanzados por tipo de gasto, área funcional o nivel institucional.
- Despliegue en servidores institucionales o en la nube para acceso multiusuario.

9.3 Recomendaciones para productivización

Para convertir el dashboard en una herramienta institucional, se recomienda:

- Validar su funcionamiento con datos reales en entornos controlados.
- Documentar cada módulo del código con comentarios claros y pruebas unitarias.
- Establecer un sistema de control de versiones para facilitar su mantenimiento.
- Diseñar una interfaz de usuario adaptable a dispositivos móviles y accesible según estándares públicos.
- Formar a los usuarios finales en el uso de la herramienta, especialmente en entornos administrativos.

9.4 Evaluación del módulo predictivo

La inclusión del módulo de predicción presupuestaria aporta un valor añadido al visor, al permitir no solo analizar y comparar datos históricos, sino también proyectar escenarios futuros.

Aunque se ha optado por un modelo de regresión lineal por su simplicidad e interpretabilidad, se ha estructurado el código para permitir la incorporación de modelos más complejos como Random Forest o XGBoost en futuras versiones.

El modelo ha sido validado con datos reales, y aunque no se ha realizado una evaluación exhaustiva de métricas como R^2 o MAE en esta versión, se ha comprobado que las predicciones generadas son coherentes con las tendencias observadas en los datos históricos.

Esta funcionalidad refuerza el carácter aplicado del TFM y demuestra la capacidad del visor para evolucionar hacia una herramienta de apoyo a la planificación financiera pública.

10. Bibliografía

10.1 Fuentes de datos

- Ministerio de Hacienda y Función Pública – Portal de Presupuestos Generales del Estado
- Portal de Datos Abiertos del Gobierno de España
- Normativa presupuestaria vigente (Ley de Presupuestos Generales)

10.2 Referencias técnicas y académicas

- McKinney, W. (2017). *Python for Data Analysis*. O'Reilly Media.
- Streamlit Documentation – <https://docs.streamlit.io>
- Plotly Express Documentation – <https://plotly.com/python/plotly-express>
- Guía de presupuestos públicos – Instituto Nacional de Administración Pública
- PEP8 – Guía de estilo para código Python

11. Anexos

Código fuente

Se incluye el código completo del visor presupuestario, estructurado en módulos:

- Carga y validación de archivos
- Limpieza y transformación de columnas
- Agrupación dinámica
- Visualización interactiva
- Comparación entre ejercicios

Cada función está documentada y comentada para facilitar su comprensión y reutilización.

Capturas del dashboard

Se adjuntan imágenes que muestran:

- La portada institucional del dashboard
- La selección progresiva de parámetros
- La visualización de tablas y gráficos
- La comparación entre ejercicios
- Los mensajes de validación y ayuda contextual

Detalles técnicos adicionales

- Estructura de carpetas utilizada para organizar los presupuestos
- Lógica de navegación por series, niveles y documentos
- Validaciones aplicadas para evitar errores de carga
- Consideraciones sobre compatibilidad entre formatos (.csv y .py)

