

17_Casos prácticos de migración desde SAS

June 18, 2025

Creado por:

Isabel Maniega

1 Casos prácticos de migración desde SAS

El DATA Step en SAS y pandas en Python son herramientas poderosas para el manejo y transformación de datos, pero difieren significativamente en su estructura y enfoque. DATA Step es un entorno de programación completo dentro de SAS, mientras que pandas es una biblioteca de Python. DATA Step se centra en la lectura, escritura y manipulación de datos, mientras que pandas proporciona estructuras de datos optimizadas (como DataFrames) y funciones para análisis exploratorio y manipulación.

Estructuras de Datos:

- **DATA Step:** Utiliza estructuras de datos como datasets (tablas) y variables. El enfoque principal es la manipulación de datos en memoria, con un énfasis en la lectura y escritura de archivos.
- **pandas:** La estructura central es el DataFrame, que es una tabla bidimensional con filas y columnas etiquetadas, similar a una hoja de cálculo o tabla SQL. También incluye Series, que son unidimensionales. pandas se destaca por su capacidad para manejar datos heterogéneos (diferentes tipos de datos en la misma tabla) y su facilidad de uso con operaciones como indexado, filtrado y agrupación.

Procesos:

- **DATA Step:** Es un entorno de programación que implica un ciclo de lectura, procesamiento y escritura de datos. Ofrece una variedad de funciones para transformación, cálculo y manipulación de datos.
- **pandas:** Se basa en funciones y métodos aplicables a objetos DataFrame y Series. Ofrece una sintaxis más intuitiva y flexible para manipulación de datos, incluyendo funciones para limpieza, transformación, agregación, y análisis exploratorio.

Comparación detallada:	Característica	DATA Step	pandas
Entorno	SAS (lenguaje y entorno completo)	Biblioteca de Python	
Estructura de datos	Datasets, variables	DataFrame, Series	
Enfoque principal	Lectura, escritura, manipulación de datos	Análisis exploratorio, manipulación de datos, transformación	
Flexibilidad	Mayor control, pero puede ser más verboso	Mayor flexibilidad y sintaxis más intuitiva	
Velocidad	Puede ser más rápido en ciertas operaciones con archivos grandes	Puede ser más rápido en operaciones complejas con datos en memoria, pero puede verse limitado	

por la memoria RAM | | Facilidad de uso | Curva de aprendizaje más pronunciada | Curva de aprendizaje más suave para usuarios familiarizados con Python | | Comunidad y soporte | Amplia comunidad SAS, documentación extensa | Amplia comunidad Python, gran variedad de recursos y tutoriales |

En resumen:

- Si ya estás trabajando en el entorno SAS, DATA Step es una opción natural.
- Si buscas una herramienta flexible y potente para manipulación y análisis de datos en Python, pandas es una excelente elección.
- pandas destaca por su facilidad de uso, velocidad en ciertas operaciones y una sintaxis más intuitiva, especialmente para usuarios familiarizados con Python.
- DATA Step ofrece mayor control sobre el proceso de lectura y escritura de datos, especialmente con archivos grandes, y es una opción robusta dentro de SAS.

Creado por:

Isabel Maniega