

Premios a la reutilización 2025

Santiago Mota | Presentación

2025-09-04



Santiago Mota Herce

Premios a la
reutilización
2025

Santiago
Mota |
Presentación



- Físico. MBA y Máster BI y BigData (2013).
- M&A. CEO empresa náutica I+D+I.
- Consultor freelance. Vodafone, Teradata, Bankia, Repsol, BCG, Orange.
- [LinkedIn](#)

Premios a la reutilización 2025

Premios a la
reutilización
2025

Santiago
Mota |
Presentación

- Premios a la reutilización 2025 [Link](#)
- Presentación [Link](#)
- Bases [Link](#) y [Documento](#)
- Categoría: Propuestas de mejora de la calidad del Portal de Datos Abiertos.

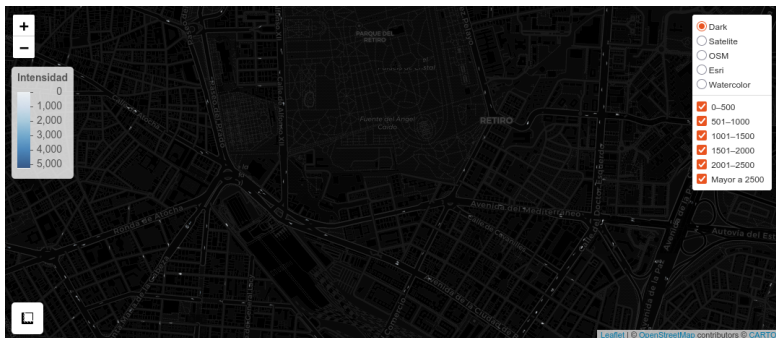
- Incorporar códigos de ejemplo en Python y R.
- Exportar datos a parquet y usar duckDB.
- Abrir una página de [Github](#).
- Incorporar datos en [Kaggle](#) (crear un usuario para el ayuntamiento).

Las dos primeras son internas y las dos segundas son externas.

- Propuesta: Incorporar en la pagina web ejemplos de código para el uso de los datos publicados tanto en Python como en R.
- Objetivos: Facilitar el tratamiento de los datos publicados.

Incorporar códigos de ejemplo en Python y R II

- Implementación: Se aportan varios ejemplos en el repositorio [Link](#) y el repositorio [Link](#). El primero son ejemplos relativamente sencillos y en el segundo se abordan problemas de mas profundidad, en el especial el del tráfico de Madrid.



Exportar datos a parquet y usar duckDB I

Premios a la
reutilización
2025

Santiago
Mota |
Presentación

- Propuesta: La incorporación de datos en formato parquet está convirtiéndose poco a poco en un estándar que sustituye a los csv.
- Objetivos: La principal ventaja es que la compresión de esta formato reduce el tamaño de los datos. Además permite el uso de herramientas como duckDB que facilitan trabajar con datos superiores a la memoria del ordenador.

Exportar datos a parquet y usar duckDB II

Premios a la
reutilización
2025

Santiago
Mota |
Presentación

- Implementación: Se ha desarrollado un código en el repositorio [Link](#) para recorrer la página web y transformar los ficheros csv a parquet.
- **`madrid_csv_to_parquet.py` para convertir todos los csv a parquet**

Abrir una página de Github

Premios a la
reutilización
2025

Santiago
Mota |
Presentación

- Propuesta: Crear una página propia en Github de Datos Abiertos del Ayuntamiento de Madrid.
- Objetivos: Facilitar el trabajo con los datos abiertos a través del control de versiones.
- Implementación: No debería llevar más de uno o dos días crear la página.

Incorporar datos en Kaggle

Premios a la
reutilización
2025

Santiago
Mota |
Presentación

- Propuesta: Crear una página propia en Kaggle de Datos Abiertos del Ayuntamiento de Madrid.
- Objetivos: Promocionar el uso de los datos abiertos del Ayuntamiento en una plataforma de uso internacional con mas de 200.000 usuarios.
- Implementación: No debería llevar mas de uno o dos días crear la página.

- El premio del millón de dólares de Netflix [Link con suscripción](#).
- Mas de 200.000 usuarios en todo el mundo.
- Zona de test para los algoritmos mas avanzados.
- Adquirido en 2017 por Google ([Wikipedia Kaggle](#)).

- Las mejoras están dirigidas a ampliar el uso de los datos abiertos del Ayuntamiento de Madrid.
- Se proponen cambios en la propia web y la creación de páginas en otras.
- Como punto negativo, la posibilidad de perder algo de control sobre el uso.
- Son bastante sencillas de implementar y se aportan varios ejemplos funcionales en las internas.

- Santiago Mota [LinkedIn](#)
- santiago_mota@yahoo.es
- Github Santiago Mota
- Esta presentación online
- Repositorio Github de esta presentación

