

Primer Ejercicio: Gestión de Dashboards

Diseñar un modelo simple, práctico y realista para encontrar dashboards duplicados en Tableau. El objetivo es optimizar el uso de recursos, evitar confusiones de negocio y mejorar la confianza general en los datos.

Modelo de Datos Propuesto

	A ▼	B	C
1	Campo	Descripción	Fuente
2	dashboard_id	Identificador único del dashboard	Metadata API
3	dashboard_name	Nombre del dashboard	Metadata API
4	project_name	Proyecto o carpeta donde está publicado	Metadata API
5	owner	Propietario o área responsable	Metadata API
6	creation_date	Fecha de creación	Metadata API
7	last_modified_date	Fecha de última modificación	Metadata API
8	tags	Etiquetas o categorías (si las hay)	Metadata API
9	data_sources	Lista de data sources utilizados (incluyendo los embebidos)	Metadata API
10	datasets_consultados	Lista de datasets consultados (ejemplo: ["dbo.persona", "dbo.cuenta"])	Metadata API + lógica de extracción
11	fields_used	Lista de KPIs o métricas que utiliza el dashboard	Metadata API
12	refresh_frequency	Frecuencia de actualización	Metadata API
13	total_views	Cantidad total de visualizaciones	Admin Insights
14	last_access_date	Fecha del último acceso	Admin Insights
15	user_list	Lista de usuarios que accedieron	Admin Insights
16			

Cómo se ve la tabla con datos de ejemplo

A	B	C	D	E	F	G	H	I	J	K
Dashboard A	Dashboard B	Similaridad Total	Coincidencia KPIs	Coincidencia Data Sources	Coincidencia Owner	Coincidencia Tags	Datasets Consultados	Vistas A	Vistas B	Estado de Revisión
DB001	DB045	85%	80%	100%	100%	50%	["dbo.persona", "dbo.cuenta"]	350	220	Pendiente
DB022	DB023	92%	90%	100%	100%	80%	["dbo.ordenes", "dbo.clientes"]	500	480	Revisado
DB011	DB099	78%	70%	100%	80%	60%	["dbo.transacciones"]	120	130	Pendiente

Además de este resumen, podríamos derivar una tabla adicional de "candidatos a eliminar", dónde quedaría el dashboard con menor cantidad de visualizaciones en los casos de duplicados.

Cómo detectamos duplicados

El sistema compara varios aspectos de los dashboards y les asigna un peso según su importancia:

- KPIs utilizados (50% del peso total)
- Data Sources (25%)
- Owner o responsable (15%)
- Tags o categorías (10%)

Se calcula el porcentaje de coincidencia en cada aspecto (por ejemplo, usando Jaccard Similarity en las listas de KPIs) y se pondera para obtener un score final. Si el score supera el 80%, lo podemos considerar como potencial duplicado.

Proceso de Revisión y Validación

Los casos detectados como posibles duplicados no se eliminan automáticamente. Primero:

- Se notifica a los owners responsables vía email.
- Los owners acceden al tablero de monitoreo para revisar el caso.
- Si es necesario, participan también responsables de negocio y el equipo de Data Governance.
- Se confirma, descarta o justifica la duplicación.
- Todo queda documentado para mantener la trazabilidad.

Flujo General de Trabajo

1. Recolectamos la metadata automáticamente desde Tableau esto con la api (Metadata API + Admin Insights).
2. Ejecutamos el algoritmo de comparación de dashboards.
3. Priorizamos los casos en función del uso (vistas totales) y la antigüedad.
4. Los owners revisan los casos propuestos.
5. Se decide consolidar, eliminar o mantener cada caso.
6. El monitoreo queda activo de forma continua.

Fuentes de Metadata Usadas

- Tableau Metadata API: información técnica de los dashboards.
 - Admin Insights (Server/Cloud): datos de uso y acceso.
 - Who Has Seen This View: Si requerimos estos datos debemos descargarlo manualmente.
-

Ideas para mejorar la Discoverabilidad

- Completar siempre la descripción funcional de cada dashboard.
- Definir categorías y etiquetas por dominio de negocio.
- Aplicar convenciones de nombres claros y consistentes.
- Limpiar regularmente dashboards obsoletos según uso.
- Integrar la metadata de Tableau en un catálogo central de datos.
- Complementar a futuro con mecanismos dinámicos de enriquecimiento automático de metadata, como machine learning, NLP o data lineage.

Propuesta de Visualizaciones del Tablero de Monitoreo

1. Resumen general de indicadores (dashboards totales, duplicados detectados, pendientes de revisión, eliminados, etc.).

2. Gráfico de barras: owners con mayor cantidad de duplicados detectados.
3. Gráfico de torta: distribución por estado de revisión (pendiente, revisado, consolidado).
4. Gráfico de dispersión: Similarity Score vs Total Views, para priorizar revisiones de alto impacto.
5. Tabla de detalle con el listado de duplicados detectados.
6. Sección especial: Candidatos a eliminar (pares de dashboards donde el duplicado con menor cantidad de vistas queda sugerido para consolidación o baja).

Este tablero permite a los equipos de Data Governance monitorear el proceso de limpieza de dashboards de forma clara, visual y continua.