

Caso Práctico - Análisis de Datos de Nómina de Funcionarios Públicos

Guía de Actividades

Objetivo

Aprender a extraer, transformar y visualizar datos de la nómina de funcionarios del portal de datos abiertos de la SFP utilizando un proceso **ETL** con **Pentaho Kettle**, almacenamiento en **PostgreSQL** y reportes interactivos en **Power BI**.

1. Arquitectura de Datos del Proceso ETL

📌 Fuente de Datos:

- Datos descargados desde el **Portal de Datos Abiertos** en formato **CSV o JSON**.

📌 Proceso ETL con Pentaho Kettle:

1. Extracción:

- Leer archivos CSV desde una carpeta específica.
- Aplicar validaciones básicas (fechas, formatos numéricos).

2. Transformación:

- Convertir campos de tipo texto a tipos adecuados (fechas, números).
- Estandarizar nombres de columnas.
- Calcular **edad** a partir de *fecha_nacimiento*.
- Clasificar **antigüedad** laboral a partir de *anho_ingreso*.

3. Carga a PostgreSQL:

- Crear y poblar la tabla **bigdata.poc.reportes_funcionarios**.
- Indexar los campos clave (*documento*, *entidad*, *anho*, *mes*).

📌 Visualización con Power BI:

- Conectar Power BI a PostgreSQL.
- Crear modelos de datos y medidas.

2. Definición de Métricas e Indicadores

◆ Indicadores Financieros:

- **Total presupuestado y total devengado** por entidad y nivel.
- **Porcentaje de ejecución presupuestaria** (*devengado/presupuestado*).
- **Distribución del gasto** por tipo de objeto del gasto.

◆ Indicadores de Personal:

- **Cantidad total de funcionarios** por nivel y entidad.
- **Distribución por género (M/F)**.
- **Promedio de edad** y distribución por rangos de edad.
- **Cantidad de funcionarios con discapacidad** y tipo de discapacidad.
- **Antigüedad promedio** del personal por nivel y entidad.

◆ Indicadores Organizacionales:

- **Cantidad de funcionarios nombrados vs. contratados**.
- **Cantidad de funcionarios por categoría laboral**.
- **Cantidad de funcionarios por OEE y nivel jerárquico**.

3. Diseño de Tablero en Power BI

📌 Páginas y Secciones del Dashboard:

- **Resumen Ejecutivo** → KPIs principales (Visión General)
- **Análisis Financiero** → Comparación entre presupuestado y devengado.
- **Análisis de Personal** → Distribución por Edad, género, antigüedad.
- **Comparación por Entidades** → Personal y gasto por entidad y OEE.

📌 Gráficos Recomendados:

📊 Indicadores Generales:

- **Tarjetas KPI** → Total presupuestado, devengado y % ejecución.
- **Matriz** → Presupuesto y gasto por entidad.

📊 Gasto Público:

- **Gráfico de columnas** → Comparación presupuestado vs. devengado.
- **Mapa de calor** → Concentración del gasto por entidad.

📊 Distribución del Personal:

- **Gráfico de torta** → Género.

-
- **Histograma** → Edad.
 - **Mapa geográfico** → Distribución por región.



Análisis de Categoría Laboral:

- **Treemap** → Categorías con más funcionarios.
 - **Gráfico de dispersión** → Antigüedad vs. sueldo.
-

4. Publicación y Análisis de Resultados

- ✓ Publicar el tablero en **Power BI Service**.
 - ✓ Configurar filtros interactivos.
 - ✓ Compartir con los usuarios clave.
 - ✓ Analizar patrones de gasto y estructura laboral.
-

Conclusión:

Este proyecto permitirá a los estudiantes **aplicar conocimientos básicos de ETL, almacenamiento y visualización** en un caso real.