

# Análisis Detallado de Notebooks en ETL\_Workshop\_02

ETL\_grammy.ipynb 

## Introducción y Objetivos

Este notebook está diseñado para sumergirse en el análisis de los datos de los premios Grammy. Su misión principal es procesar y preparar estos datos para futuros análisis, mejorando así la accesibilidad y la manipulación de la información detallada sobre premios y nominaciones. Con un enfoque centrado en los Grammy, se establece una base sólida para investigaciones o aplicaciones analíticas futuras.

## Importación de Librerías y Configuración

Aquí se destacan las librerías esenciales como pandas, para la manipulación de datos; psycopg2, para conectar con bases de datos PostgreSQL; y matplotlib, para la visualización de datos. La configuración inicial incluye la importación de parámetros esenciales desde un archivo externo, asegurando que el entorno de trabajo esté perfectamente preparado para los procesos ETL, incluyendo rutas y conexiones a la base de datos.

## Proceso ETL

### Extracción

Definimos rutas de origen y destino para manejar archivos de datos, garantizando su integridad durante la transferencia.

### Transformación

Transformamos archivos CSV a Excel, monitoreando cada operación para asegurar la ejecución correcta del proceso y la integridad de los datos.

### Carga

Preparamos los datos para su integración en sistemas de almacenamiento de datos o análisis más complejos. Aunque los detalles específicos de la carga no se muestran, la estructuración sugiere que los datos están listos para su uso futuro.

## ETL\_spotify.ipynb

### Introducción y Objetivos

Centrado en el manejo de datos de Spotify, este notebook tiene como objetivo transformar datos musicales en información relevante para análisis musical. Subraya la importancia de procesar eficientemente los datos para descubrir insights útiles para la industria musical.

### Importación de Módulos y Configuración

Uso de pandas para la manipulación de datos y matplotlib para visualizaciones, junto con otras librerías estándar para la gestión de archivos y directorios. La configuración inicial prepara el entorno configurando rutas de archivos y parámetros esenciales.

## Proceso ETL

### Extracción

Establecemos rutas de archivos y manejamos la copia de archivos necesarios para el procesamiento, asegurando la disponibilidad de los datos.

## Transformación

Transformación y limpieza de datos, con la creación de visualizaciones que ayudan a entender y verificar la calidad de los datos procesados.

## Carga

Preparamos los datos para su integración en sistemas de almacenamiento de datos o para su uso en herramientas de análisis, aunque los detalles completos no se proporcionan en el notebook analizado.

## ETL\_merge.ipynb

### Objetivo

Este notebook se enfoca en la fusión y limpieza de datos provenientes de diversas fuentes, diseñado específicamente para consolidar estos datos en un formato coherente y optimizado para análisis. Es crucial para la integración de bases de datos y archivos CSV, asegurando que la información consolidada sea precisa y útil para el análisis o almacenamiento de datos.

## Importación de Módulos y Configuración

Utilización de librerías como pandas para manipulación de datos y psycopg2 para operaciones de base de datos, reflejando la interacción continua con bases de datos PostgreSQL. Configuración de rutas desde un archivo externo para manejar la ubicación de los archivos de datos de manera efectiva.

## Proceso ETL


### Extracción

Carga de datos desde archivos CSV en DataFrames, preparando el escenario para su manipulación y fusión.

## Transformación

Fusión de datos de múltiples fuentes, limpieza de datos duplicados, y preparación de los datos para análisis más complejos, mostrando la habilidad del notebook para manejar y consolidar información de manera eficiente.

\*C

arga\*\* 

Aunque el proceso de carga no se detalla, los datos preparados están listos para su integración en sistemas de almacenamiento de datos o para análisis adicionales.

## Creación del Archivo "Blmerge.pbix"

Después de realizar las operaciones de extracción, transformación y carga detalladas en los notebooks anteriores, todo el trabajo culminó en la creación del archivo "Blmerge.pbix". Este archivo es un informe de Power BI diseñado para la visualización y análisis de datos consolidados. Power BI es una herramienta de inteligencia de negocios que permite crear reportes dinámicos y dashboards interactivos.

## Tipo de Datos Manejados

El archivo "Blmerge.pbix" integra diversos conjuntos de datos que incluyen:

- **Datos de Premios Grammy:** Información detallada sobre nominaciones y ganadores, categorías de premios, y datos históricos relevantes.
- **Datos de Spotify:** Análisis de tendencias musicales, datos de artistas, álbumes, y pistas, incluyendo métricas como popularidad y reproducciones.
- **Datos Consolidados de Diversas Fuentes:** Información que ha sido fusionada y limpiada para asegurar consistencia y precisión, permitiendo análisis cruzados entre diferentes tipos de datos relacionados con la industria musical.

## Utilidad del Archivo

El "Blmerge.pbix" sirve como una herramienta crítica para tomar decisiones basadas en datos, proporcionando insights comprensivos sobre las tendencias de la industria musical y el impacto de los premios Grammy. Los usuarios pueden interactuar con los datos mediante filtros y visualizaciones, lo que facilita la exploración y el análisis detallado de la información.