

Manual de Usuario para el Proceso ETL de la Red Eléctrica

Grupo de Control y Procesamiento Digital de Señales
Universidad Nacional de Colombia
CHEC

April 28, 2025

Contents

1	Descripción General	2
2	Funcionalidades Principales	2
3	Bibliotecas Necesarias	2
4	Archivos Necesarios	2
5	Instrucciones de Uso	3
5.1	Preparación del Entorno	3
5.2	Ejecución del Código	3
5.3	Pasos Detallados	3
6	Consideraciones Adicionales	3

1 Descripción General

Este código de Python está diseñado para realizar un proceso de Extracción, Transformación y Carga (ETL) de datos relacionados con la red eléctrica. El objetivo principal es consolidar y limpiar datos provenientes de diferentes fuentes, preparándolos para análisis posteriores. El proceso incluye la unión de archivos, eliminación de duplicados, transformación de datos geográficos y generación de resúmenes.

2 Funcionalidades Principales

- **Unión y Limpieza de Datos:** Combina múltiples archivos CSV con datos de la red eléctrica, eliminando duplicados y columnas no necesarias.
- **Transformación de Coordenadas:** Convierte coordenadas de un sistema de referencia a otro (EPSG:3115 a EPSG:4326).
- **Asignación de Municipios y Departamentos:** Asigna el municipio y departamento más cercanos a cada punto geográfico.
- **Resúmenes y Reportes:** Genera archivos Excel con resúmenes de los datos procesados.
- **Filtrado y Actualización de Datos:** Filtra y actualiza datos específicos según condiciones predefinidas.

3 Bibliotecas Necesarias

Para ejecutar este código, se requieren las siguientes bibliotecas de Python:

- **pandas:** Utilizado para manipulación y análisis de datos.
- **numpy:** Utilizado para operaciones numéricas.
- **pyproj:** Utilizado para transformaciones de coordenadas geográficas.
- **scipy:** Utilizado para operaciones de ciencia de datos, como el árbol KD.
- **openpyxl:** Utilizado para escribir archivos Excel.
- **matplotlib:** Utilizado para visualización de datos.
- **requests:** Utilizado para solicitudes HTTP.
- **google.colab:** Utilizado para interactuar con Google Colab.

4 Archivos Necesarios

Los siguientes archivos deben existir y ser accesibles para que el código funcione correctamente:

- **Archivos CSV:** Datos de la red eléctrica con prefijos específicos (e.g., APOYOS, EVENTOS, REDMT, SWITCHES, TRAFOS).
- **Archivos Excel:** Datos adicionales y parámetros (e.g., Parámetros-varios.xlsx, Conductores.xlsx).
- **Archivo de Clasificador Geográfico:** Contiene información geográfica para la transformación de coordenadas.
- **Google Drive:** Acceso a una carpeta compartida en Google Drive para leer y escribir archivos.

5 Instrucciones de Uso

5.1 Preparación del Entorno

1. **Instalar Bibliotecas:** Asegúrate de tener todas las bibliotecas necesarias instaladas en tu entorno de Python. Puedes instalarlas usando pip:

```
pip install pandas numpy pyproj scipy openpyxl matplotlib requests google-colab
```

2. **Configurar Google Colab:** Si estás utilizando Google Colab, asegúrate de montar tu Google Drive:

```
from google.colab import drive
drive.mount('/content/drive')
```

5.2 Ejecución del Código

1. **Configurar Rutas:** Asegúrate de que las rutas de los archivos de entrada y salida estén correctamente configuradas en el código. 2. **Ejecutar el Script:** Ejecuta el script completo. El proceso se dividirá en varias secciones, cada una correspondiente a una funcionalidad principal. 3. **Verificar Resultados:** Revisa los archivos generados en la carpeta de salida. Los archivos procesados y los resúmenes se guardarán en formato CSV y Excel.

5.3 Pasos Detallados

1. **Unión y Limpieza de Datos:** El script buscará archivos CSV con los prefijos especificados, los unirá y eliminará duplicados.
2. **Transformación de Coordenadas:** Las coordenadas en los archivos se transformarán de EPSG:3115 a EPSG:4326.
3. **Asignación de Municipios y Departamentos:** Se asignará el municipio y departamento más cercanos a cada punto geográfico.
4. **Resúmenes y Reportes:** Se generarán archivos Excel con resúmenes de los datos procesados.
5. **Filtrado y Actualización de Datos:** Los datos se filtrarán y actualizarán según las condiciones predefinidas.

6 Consideraciones Adicionales

- **Rutas de Archivos:** Asegúrate de que las rutas de los archivos de entrada y salida sean correctas y accesibles.
- **Permisos de Google Drive:** Asegúrate de tener los permisos necesarios para leer y escribir en la carpeta compartida de Google Drive.
- **Backup de Datos:** Realiza copias de seguridad de los datos originales antes de ejecutar el script.
- **Dependencias de Bibliotecas:** Asegúrate de que todas las bibliotecas estén actualizadas y funcionen correctamente en tu entorno.