**Documentación del proceso (ETL).**



**1. Introducción.**

Este documento describe el proceso de ETL para la carga de datos de gastos e ingresos de clínicas médicas desde archivos almacenados en carpetas asignadas, su transformación en DataFrames de Python, generación de índices, agrupación de datos y carga en una base de datos SQL Server. Además, se describe la conexión de SQL Server con Power BI para análisis y visualización de los datos.

**2. Proceso ETL en Python.**

**2.1. Extracción de Datos:** Los datos se extraen de archivos ubicados en carpetas predefinidas utilizando Python. Se utilizan librerías como “os” para la navegación de directorios y “pandas” para la lectura de archivos xlsx.

**2.2. Transformación de Datos:** Se generan DataFrames separados para los archivos de gastos e ingresos. Cada DataFrame se indexa y se les asigna un identificador único. Luego, se realiza la agrupación de datos para generar un índice único que permita hacer referencia entre gastos e ingresos.

**2.3. Carga de Datos en SQL Server:** Los DataFrames transformados se cargan en tablas de una base de datos SQL Server. Se utiliza la librería “pyodbc” para establecer la conexión con la base de datos y “pandas” para la carga de datos.

**3. Conexión con Power BI.**

**3.1. Configuración de Conexión:** Se establece una conexión entre SQL Server y Power BI para importar los datos.

**A diagram of a computer

Description automatically generated**