**Proyecto BD**

**Integrantes:**

**Valentina Valencia Quejada**

**Yuli Andrea Goez**

**Dylan David Marmolejo**

**Docente: Jaime Gómez**

**Facultad de Ingeniería**

**Politécnico Colombiano Jaime Isaza Cadavid**

**2024**

Tabla de Contenido

[**Objetivo** 3](#_Toc177232184)

[**Descripción** 3](#_Toc177232185)

[**Consideraciones clave** 3](#_Toc177232186)

[**Diccionario de datos** 3](#_Toc177232187)

# **Objetivo**

Migrar y normalizar los datos de un dataset en CSV a una base de datos SQL, eliminando nulos, separando los valores múltiples en columnas, y realizando la indexación para optimizar las consultas.

# **Descripción**

Vamos a cargar los datos en un DataFrame utilizando **Pandas**, realizar limpieza de datos y normalización, y posteriormente migrarlos a una base de datos en SQL Server. El proceso incluye los siguientes pasos:

* Ver los parámetros del DataFrame.
* Normalizar columnas con múltiples valores.
* Eliminar registros nulos.
* Crear índices en la base de datos SQL para optimizar consultas.

# **Consideraciones clave**

* Se debe usar **Pandas** para manejar el dataset.
* La migración a SQL se puede hacer utilizando **SQLAlchemy** o cualquier conector que soporte SQL Server.
* Para la normalización, cuando hay más de dos elementos separados por comas, se separarán en columnas adicionales.

# **Diccionario de datos**

|  |  |  |
| --- | --- | --- |
| **Columna** | **Tipo de dato** | **Descripción** |
| **país** | **varchar(100)** | **Nombre del país asociado al registro** |
| **código iso país** | **char(3)** | **Código ISO de 3 letras que representa el país** |
| **oficina de registro** | **varchar(100)** | **Nombre de la oficina o ciudad donde se registró la información.** |
| **grupo edad** | **varchar(50)** | **Grupo de edad de la persona (por ejemplo, adulto, adulto joven).** |
| **edad (años)** | **float** | **Edad de la persona en años.** |
| **etnia de la persona** | **varchar(100)** | **Etnia de la persona registrada, si está disponible.** |
| **estatura (cm)** | **float** | **Estatura de la persona en centímetros.** |
| **Cantidad de personas** | **int** | **Número de personas involucradas en el registro.** |
| **latitud** | **float** | **Coordenada geográfica de latitud del lugar asociado al registro.** |
| **longitud** | **float** | **Coordenada geográfica de longitud del lugar asociado al registro.** |