

**Proyecto Final**

**Profesora:**

**Natalia Herrera**

**Estudiantes:**

**Gary Carballo Barrantes**

**Marco Fonseca León**

**André Montoya Fernández**

**LEAD University**

**2023**

Logotipo

Descripción generada automáticamente

**Introducción**

En el ámbito actual del analisis de datos, la calidad de los datos es un componente crucial para la toma de decisiones acertadas e informadas. Este proyecto se centra en proporcionar una sólida base en términos de limpieza de datos, abordando los problemas más comunes que pueden surgir al trabajar con conjuntos de datos diversos. La misión en este caso es la detección de esos problemas en con las diferentes funciones aplicadas para saber a qué nos enfrentamos.

**Desarrollo**

Tenemos dos archivos que contienen los códigos para efectuar la detección de los problemas en cuanto a calidad de los datos. El archivo **Main.py** es el archivo de principal que nos proporciona la información necesaria para cargar y analizar un archivo CSV. Lo que hace el código dentro de este archivo es lo siguiente:

1. El programa solicita al usuario el nombre y la ruta del archivo CSV, almacenando la información en un diccionario.
2. Utilizamos la función de **cargar archivo** para tomar el archivo CSV y convertir las columnas con fechas al formato adecuado.
3. Se realiza un análisis detallado del archivo cargado, incluyendo la detección de caracteres especiales, análisis de datos, verificación de columnas de fecha y evaluación de la calidad del archivo.
4. Se realizan llamadas a diversas funciones del módulo **funciones** para realizar análisis específicos, proporcionando una visión integral del contenido y la calidad del archivo.

Logotipo

Descripción generada automáticamente

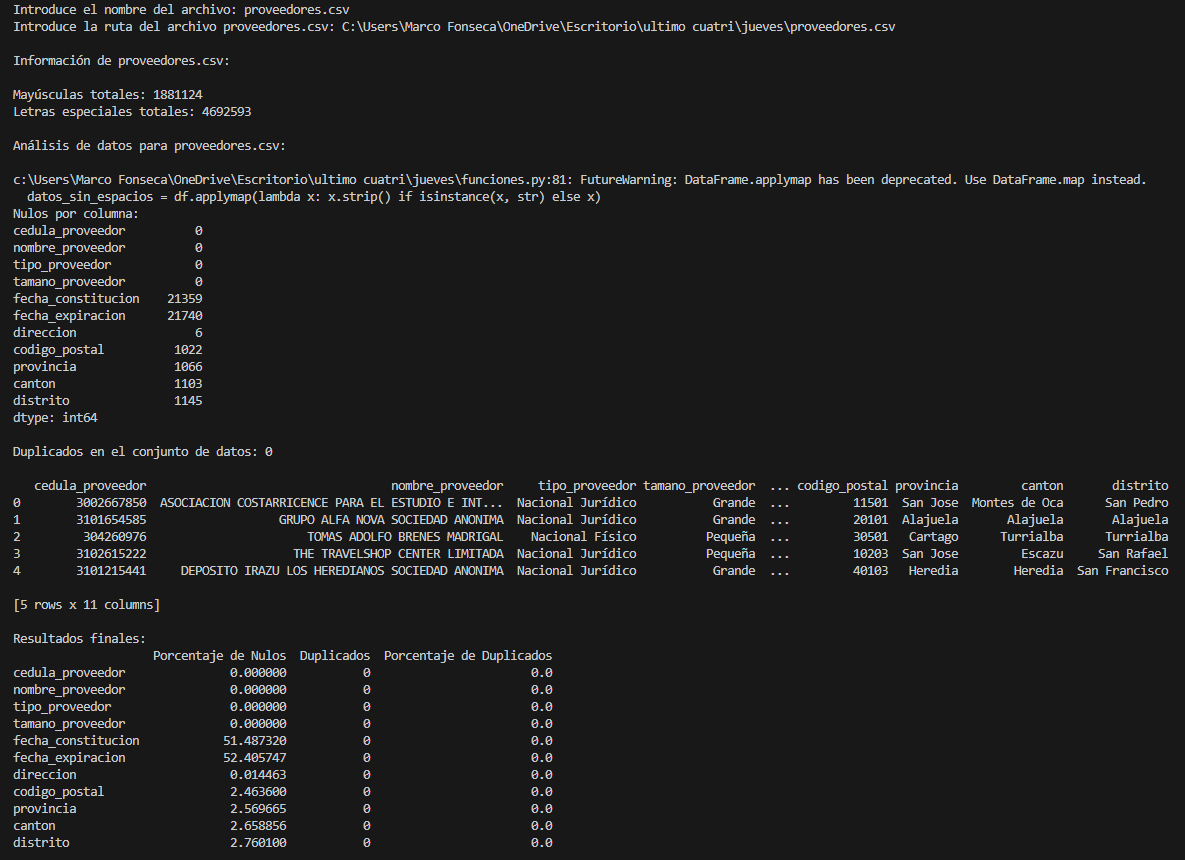
Ahora en el archivo de **funciones.py** tenemos todas las funciones aplicadas para poder realizar el análisis del archivo y por ende la detección de esos errores que nos podrían complicar nuestro trabajo más adelante.

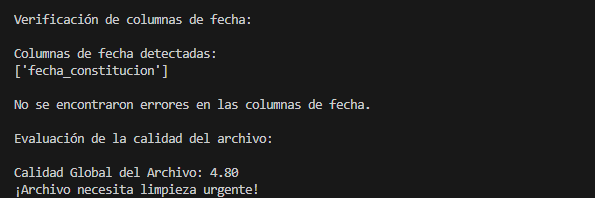
**El archivo trabaja en este orden:**

1. Cargamos el CSV y trata de convertir columnas con fechas al formato adecuado, proporcionando un DataFrame listo para el análisis.
2. Cuenta el número de mayúsculas y letras especiales en un texto, facilitando el análisis de patrones de caracteres en cadenas de texto.
3. Detecta y cuenta el número total de mayúsculas y letras especiales en columnas, ofreciendo información sobre la diversidad de caracteres presentes.
4. Realiza un análisis detallado de los datos, incluyendo el conteo de nulos, duplicados y muestra algunos datos sin espacios, proporcionando una visión general de la calidad de los datos.
5. Verifica las columnas que contienen fechas y trata de convertirlas a objetos de fecha, gestionando posibles errores de formato y rango.
6. Evalúa la calidad del archivo basándose en factores como nulos y duplicados, proporcionando un puntaje global y una evaluación cualitativa de la calidad del archivo.

Logotipo

Descripción generada automáticamente

**Resultados**



Logotipo

Descripción generada automáticamente

**Conclusión**

Después de aplicar con éxito las funciones diseñadas para la detección de factores nulos, duplicados, mayúsculas y letras especiales, logramos obtener un análisis completo de la calidad del archivo. Este proyecto no se limitó únicamente a la selección y limpieza de un conjunto de datos, sino que se enfocó en la identificación de errores.

Durante el desarrollo, nos enfrentamos a restricciones similares a las que podríamos encontrar en un entorno empresarial real. Abordar un problema concreto y trabajar dentro de ciertos límites nos permitió no solo aplicar técnicas de limpieza de datos, sino también comprender cómo gestionar desafíos y limitaciones en un contexto más práctico.

Con esta experiencia, hemos fortalecido nuestras habilidades en la detección y resolución de problemas específicos en conjuntos de datos, preparándonos para enfrentar desafíos similares en futuros proyectos. Este enfoque práctico ha enriquecido nuestro conocimiento y capacidad para abordar problemas del mundo real de manera efectiva.

Logotipo

Descripción generada automáticamente