

Avances Proyecto 1 DS

Descripción de los datos

El conjunto de datos proporcionado consta de 9378 filas y 18 columnas. A continuación se presenta un resumen de su estructura y una estrategia detallada para su limpieza:

Descripción del Conjunto de Datos

- **Número de filas:** 9378
- **Número de columnas:** 18

Estructura del conjunto de datos:

- **CODIGO:** 9355 valores no nulos
- **DISTRITO:** 9124 valores no nulos
- **DEPARTAMENTO:** 9332 valores no nulos
- **MUNICIPIO:** 9332 valores no nulos
- **ESTABLECIMIENTO:** 9332 valores no nulos
- **DIRECCION:** 9281 valores no nulos
- **TELEFONO:** 8797 valores no nulos
- **SUPERVISOR:** 9123 valores no nulos
- **DIRECTOR:** 8435 valores no nulos
- **NIVEL:** 9332 valores no nulos
- **SECTOR:** 9332 valores no nulos
- **AREA:** 9332 valores no nulos
- **STATUS:** 9332 valores no nulos
- **MODALIDAD:** 9332 valores no nulos
- **JORNADA:** 9332 valores no nulos
- **PLAN:** 9332 valores no nulos
- **DEPARTAMENTAL:** 9332 valores no nulos

Estrategias de limpieza

Variables que Necesitarán Más Operaciones de Limpieza

Basándonos en la cantidad de valores no nulos y la naturaleza de los datos, las siguientes variables probablemente requerirán más atención durante la limpieza:

- CODIGO
- DISTRITO
- MUNICIPIO
- ESTABLECIMIENTO
- DIRECCION
- TELEFONO
- SUPERVISOR
- DIRECTOR

Estrategia de Limpieza

Para asegurar la calidad de los datos, se implementará la siguiente estrategia de limpieza para las variables clave:

1. **Variable: CODIGO**
 - **Estandarización:** Asegurarse de que todos los códigos sigan un formato consistente.
 - **Eliminación de duplicados:** Remover códigos duplicados para mantener la unicidad.
2. **Variable: DISTRITO**
 - **Estandarización de nombres:** Convertir todos los nombres a minúsculas.
 - **Corrección de errores ortográficos:** Verificar y corregir posibles errores en los nombres de los distritos.
3. **Variable: DEPARTAMENTO**
 - **Estandarización de nombres:** Convertir todos los nombres a minúsculas.
 - **Corrección de errores ortográficos:** Verificar y corregir posibles errores en los nombres de los departamentos.
4. **Variable: MUNICIPIO**
 - **Estandarización de nombres:** Convertir todos los nombres a minúsculas.
 - **Corrección de errores ortográficos:** Verificar y corregir posibles errores en los nombres de los municipios.
5. **Variable: ESTABLECIMIENTO**
 - **Conversión de texto:** Convertir todo a minúsculas.
 - **Eliminación de duplicados:** Remover registros duplicados.

- **Revisión de errores ortográficos:** Corregir errores de letras en los nombres.
- 6. **Variable: DIRECCION**
 - **Estandarización de formatos:** Estandarizar el formato de las direcciones
 - **Eliminación de duplicados:** Remover registros duplicados.
- 7. **Variable: TELEFONO**
 - **Limpieza de caracteres:** Eliminar caracteres no numéricos.
 - **Consistencia en longitud:** Asegurar que todos los números de teléfono tengan una longitud consistente.
 - **Eliminación de duplicados:** Remover números de teléfono duplicados.
- 8. **Variable: SUPERVISOR**
 - **Estandarización de nombres:** Convertir todos los nombres a minúsculas.
- 9. **Variable: DIRECTOR**
 - **Estandarización de nombres:** Convertir todos los nombres a minúsculas.

Implementar esta estrategia ayudará a mejorar la calidad y la consistencia del conjunto de datos, facilitando así un análisis más preciso y confiable.