

Universidad del Valle de Guatemala

Facultad de Ingeniería

Departamento de Ciencias de la Computación

Proyecto 1: Obtención y Limpieza de los Datos

Curso: *Data Science*

Profesor: *Luis Furlán*

Libro de Códigos

Autores:

Andy Fuentes 22944

Davis Roldán 22672

Gabriel Paz 221087

Fecha: 8 de agosto de 2025

Variable	Tipo	Descripción
CODIGO	string	Código único del establecimiento, formado por depto-muni-correlativo-tipo.
CODIGO_DEPTO	int	Código numérico del departamento (extraído del CODIGO).
CODIGO_MUNI	int	Código numérico del municipio (extraído del CODIGO).
DISTRITO	string	Código del distrito educativo asignado por MINEDUC.
DEPARTAMENTO	string	Nombre del departamento donde se ubica el establecimiento.
MUNICIPIO	string	Nombre del municipio donde se ubica el establecimiento.
ESTABLECIMIENTO	string	Nombre del establecimiento educativo, estandarizado en mayúsculas.
DIRECCION	string	Dirección física del establecimiento, con espacios y formato limpiado.
TELEFONO	string	Número de teléfono del establecimiento (solo dígitos).
TELEFONO_INVALIDO	bool	Indica si el teléfono no tiene exactamente 8 dígitos (True = inválido).
SUPERVISOR	string	Nombre del supervisor asignado al establecimiento.
DIRECTOR	string	Nombre del director del establecimiento.
NIVEL	string	Nivel educativo que imparte el establecimiento.
SECTOR	string	Tipo de administración: PRIVADO (5418), OFICIAL (874), COOPERATIVA (213), MUNICIPAL (95), DESCONOCIDO (1).
SECTOR_ENCODED	int	Codificación numérica del campo SECTOR.
AREA	string	Ubicación geográfica: URBANA (5250), RURAL (1349), DESCONOCIDO (2).
AREA_ENCODED	int	Codificación numérica del campo AREA.
STATUS	string	Estado del establecimiento: ABIERTA (6600), DESCONOCIDO (1).
STATUS_ENCODED	int	Codificación numérica del campo STATUS.
MODALIDAD	string	Modalidad de enseñanza: MONOLINGUE (6391), BILINGUE (209), DESCONOCIDO (1).
MODALIDAD_ENCODED	int	Codificación numérica del campo MODALIDAD.
JORNADA	string	Jornada: DOBLE (1957), VESPERTINA (1836), MATUTINA (1687), SIN JORNADA (1001), NOCTURNA (106), DESCONOCIDO (14).
JORNADA_ENCODED	int	Codificación numérica del campo JORNADA.
PLAN	string	Plan de estudios: DIARIO(REGULAR) (4052), FIN DE SEMANA (1310), SEMIPRESENCIAL (490+), A DISTANCIA (116), VIRTUAL A DISTANCIA (52), SABATINO (5), INTERCALADO (2), DOMINICAL (2), MIXTO (1), DESCONOCIDO (1).
DEPARTAMENTAL	string	Dirección Departamental de Educación responsable del establecimiento.