



# Obtención y Limpieza de Datos de Establecimientos Educativos Diversificados en Guatemala

- Gabriela Mazariegos - 22513
- Giovanni Santos - 22523
- Santiago Pereira - 22318
- Brandon Reyes - 22992



# Objetivo de esta sección

Obtener, unificar, limpiar y preparar un conjunto de datos de establecimientos educativos a nivel diversificado en Guatemala, garantizando su calidad, consistencia y utilidad para análisis futuros.

- ✓
- ✓
- ✓





# Metodología

- Fuente de datos: Ministerio de Educación (MINEDUC)
- Formato de origen: Múltiples archivos .csv por departamento
- Herramientas utilizadas: Jupyter Notebook (Python), Pandas, Matplotlib
- Procesos realizados:
  - Análisis exploratorio inicial
  - Limpieza de nombres, jornadas, planes y sectores
  - Unificación y estandarización de datos
  - Verificación de duplicados e inconsistencias

# ESTADO CRUDO DE LOS DATOS

- Se identificaron problemas comunes como:
  - Datos nulos y registros duplicados
  - Variaciones ortográficas en campos clave (ej. nombres de jornada)
  - Inconsistencias en sectores y tipos de plan educativo

# PROCESO DE LIMPIEZA

- Limpieza de:
  - Nombres de establecimientos
  - Jornadas (estandarización)
  - Planes educativos
  - Sectores
  - Unificación en un solo dataset
  - Validación manual de registros atípicos
  -



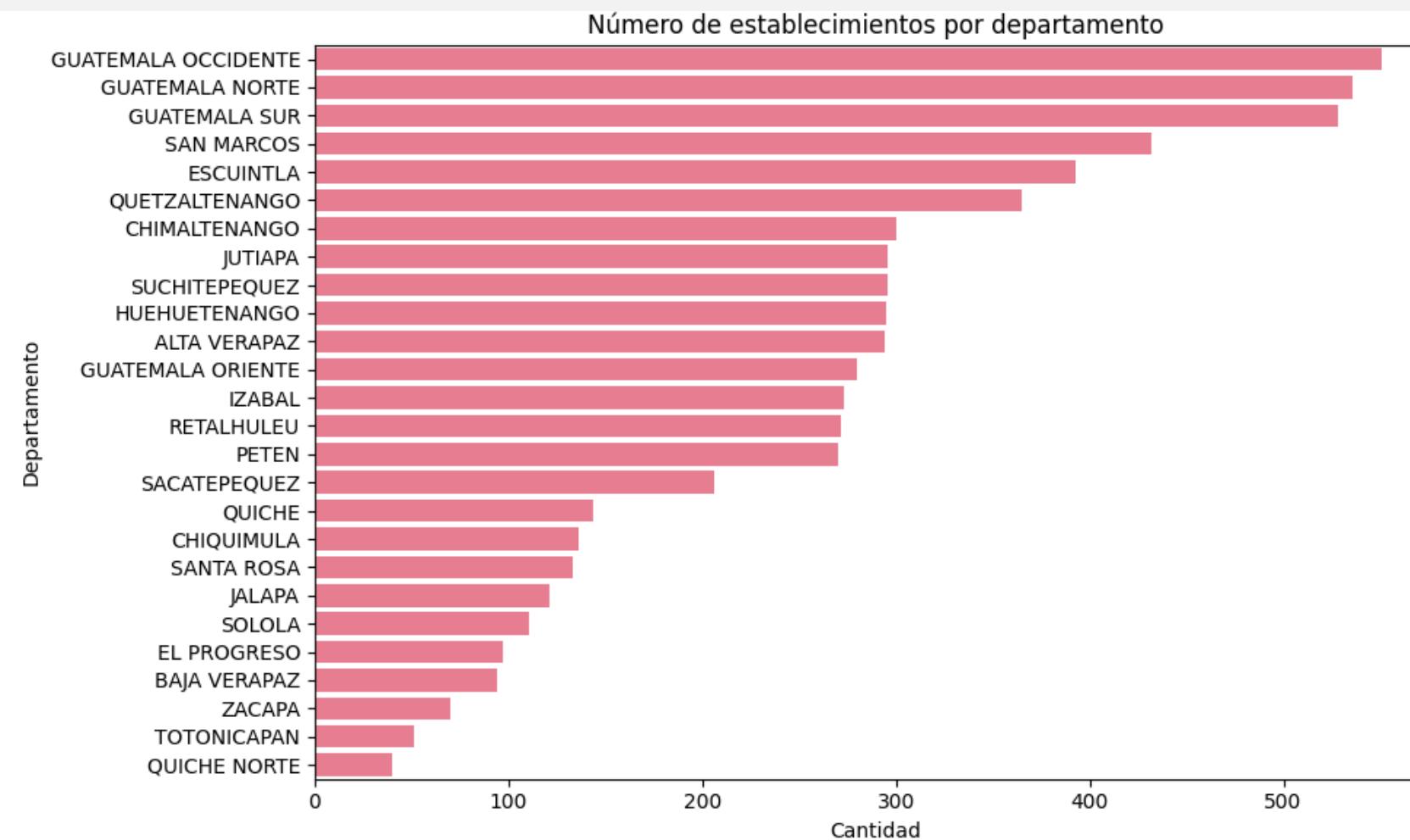
# PROCESO DE LIMPIEZA

- Limpieza de:
  - Nombres de establecimientos
  - Jornadas (estandarización)
  - Planes educativos
  - Sectores
- Unificación en un solo dataset
- Validación manual de registros atípicos



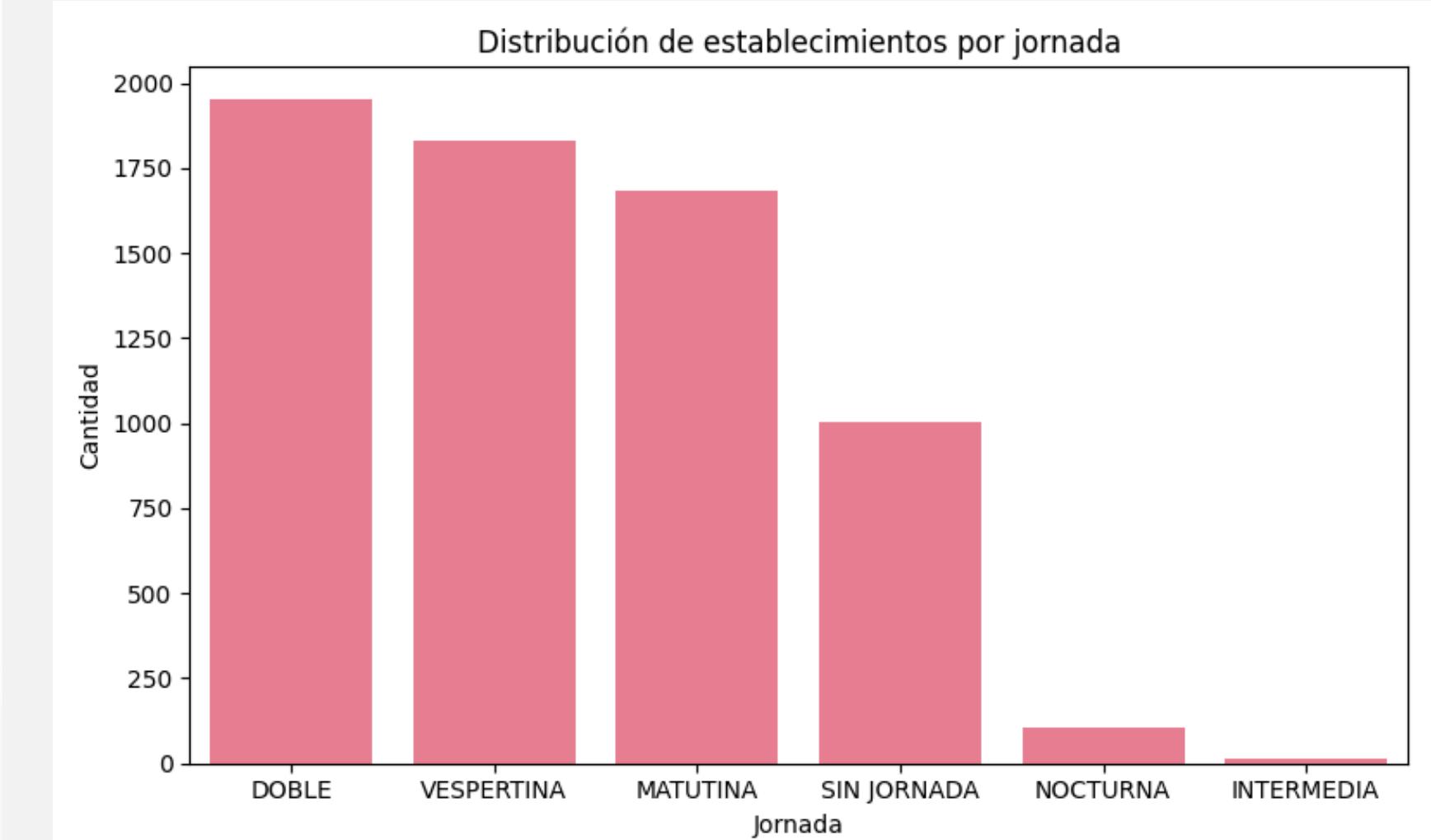
# VISUALIZACIONES DESTACADAS

## Establecimientos por Departamento



Guatemala y Ciudad Capital concentran el mayor número de establecimientos diversificados.

## Establecimientos por Jornada

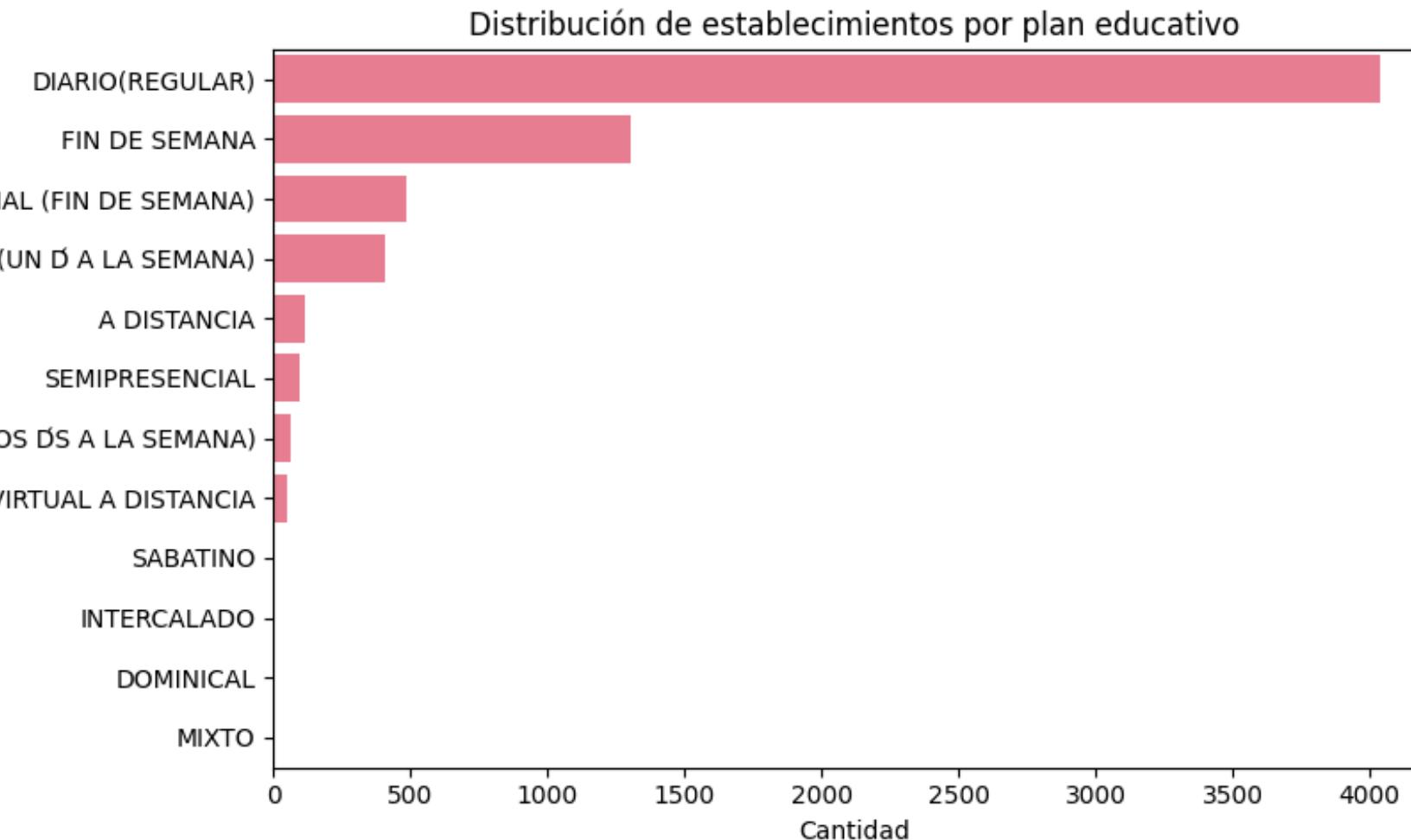


La mayoría opera en jornadas doble, vespertina y matutina. Muy pocos usan jornada intermedia o nocturna.

# VISUALIZACIONES DESTACADAS

## Establecimientos por Plan Educativo

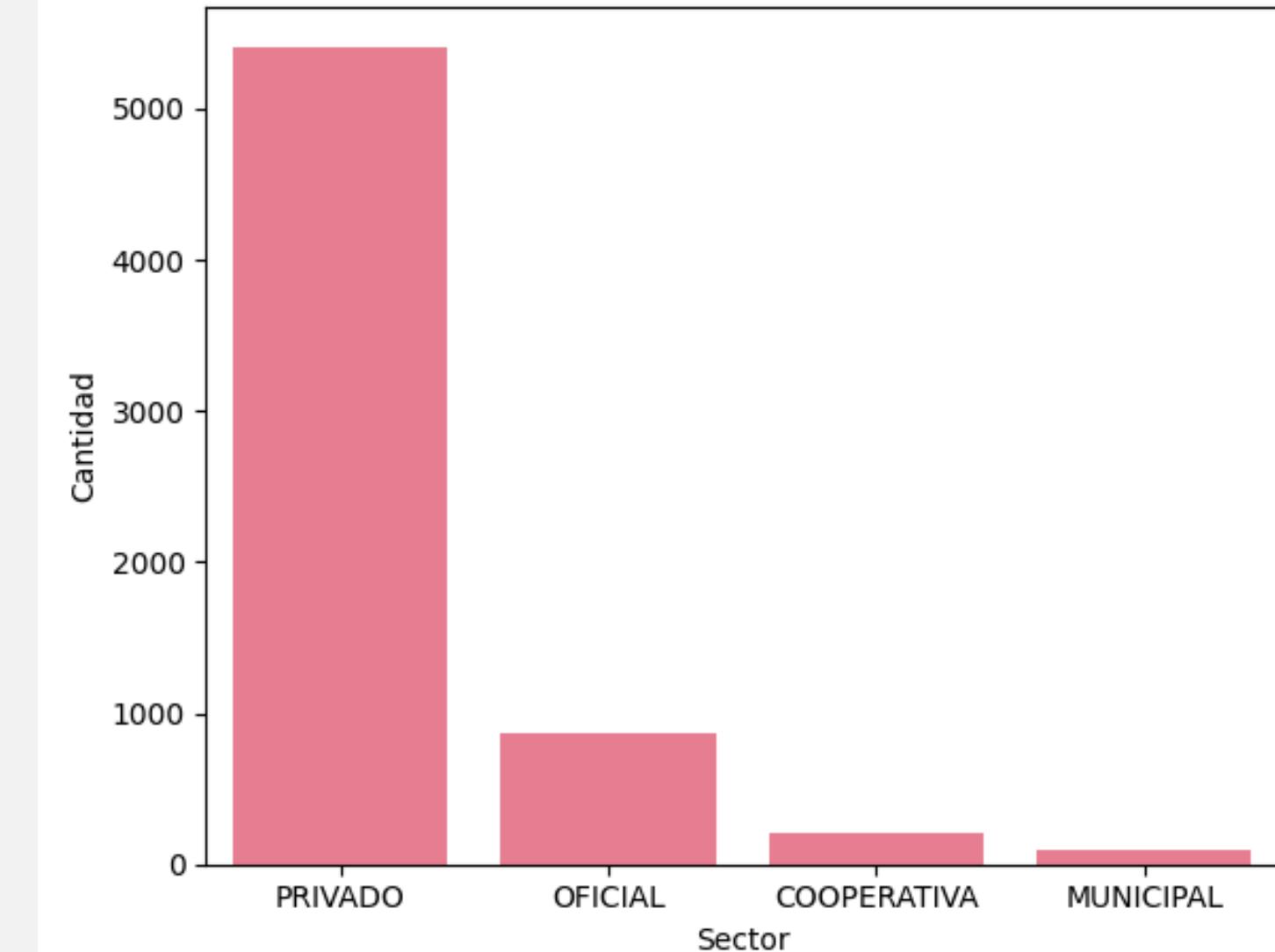
Plan



El plan Diario (Regular) domina ampliamente. Los planes a distancia o mixtos son minoritarios.

## Establecimientos por Sector

Distribución por sector (Público / Privado)



La mayoría opera en jornadas doble, vespertina y matutina. Muy pocos usan jornada intermedia o nocturna.

# CONCLUSIONES

● El proceso de limpieza es fundamental para garantizar la calidad de los datos.

● Se evidencian desequilibrios regionales y sectoriales en la oferta educativa diversificada.

● El dataset final permite realizar análisis descriptivos, predicciones y segmentaciones.