

## **Análisis Exploratorio de Datos (EDA) – Estudiantes con TEA en Cali** **(Fuente: SIMAT)**

**Fuente de datos:** Archivo con una hoja de cálculo (Hoja1), procesada mediante lectura en streaming para optimizar el análisis de hasta 50.000 registros.

### **1. Contexto del SIMAT y Definiciones Clave**

- **TEA (Trastorno del Espectro Autista):** Condición del neurodesarrollo que puede implicar necesidades de apoyo en áreas como la comunicación, la interacción social y la flexibilidad conductual.
- **SIMAT:** Sistema de Matrícula que centraliza la información oficial de los estudiantes. Su adecuada utilización es fundamental para la planificación educativa, el seguimiento a la permanencia escolar y la asignación de apoyos.
- **PIAR (Plan Individual de Ajustes Razonables):** Documento que establece los apoyos y adaptaciones requeridas por cada estudiante, en articulación con los datos registrados en SIMAT.
- **Roles y responsabilidades:**
  - *Secretaría de Educación:* Orientación y monitoreo del proceso.
  - *Instituciones Educativas Oficiales (IEO):* Liderazgo en el registro a través de rectoría y orientación.
  - *Docentes:* Implementación del PIAR y reporte de actualizaciones.
  - *Familias:* Aporte de documentación y participación activa.
- **Validaciones comunes:** Verificación de identidad del estudiante, matrícula (sede, jornada, modelo), registro de condición de discapacidad/TEA con soporte, coherencia entre PIAR y SIMAT, y control de registros duplicados.
- **Categorías analizadas:** Tipo de discapacidad o condición, apoyos recibidos, grado y edad, institución educativa, sede, jornada, ubicación geográfica (comuna), y variables socioeconómicas (como estrato o Sisbén), según disponibilidad.

## 2. Metodología del Análisis

- Aplicación de una normalización básica del texto y detección de menciones a TEA mediante patrones como “autis” y “TEA”.
- Identificación heurística de columnas clave: año, grado, sexo, comuna, institución, estrato y edad.
- Elaboración de tablas de distribución para totales, casos TEA y porcentajes por categoría.
- Evaluación de calidad de datos mediante el análisis del porcentaje de valores faltantes por columna.

## 3. Resultados del Análisis (Hoja1)

### Resumen general:

- **Filas analizadas:** 50.000
- **Columnas:** 62
- **Casos TEA detectados:** 0
- **Columnas potencialmente relacionadas con TEA:** tipo\_discapacidad
- **Columnas clave identificadas:** grado, genero, estrato, nombre institucion, comuna

### Columnas con mayor porcentaje de datos faltantes (Top 15):

Columna	% Faltantes
ano_inf	98.96%
metodologia	98.96%
res_mun	98.96%
estrato	98.96%
zon_alu	98.96%
fecha_nacimiento	98.96%

nac_depto	98.96%
nac_mun	98.96%
fue_recu	98.96%
mun_codigo	98.96%
tipo_discapacidad	98.96%
cap_exc	98.96%
etnia	98.96%
res	98.96%
tipo_jornada	98.96%

#### **Distribución de TEA por Grado (Top 10):**

Todos los grados analizados reportaron 0 casos de TEA. Ejemplos:

<b>Grado</b>	<b>Total Estudiantes</b>	<b>Casos TEA</b>	<b>% TEA</b>
0.0	68	0	0.0%
10.0	34	0	0.0%
11.0	23	0	0.0%

#### **Distribución por Género:**

<b>Género</b>	<b>Total</b>	<b>Casos TEA</b>	<b>% TEA</b>
F	96	0	0.0%
M	424	0	0.0%
Sin dato	49,480	0	0.0%

#### **Distribución por Estrato:**

<b>Estrato</b>	<b>Total</b>	<b>Casos TEA</b>	<b>% TEA</b>
1	154	0	0.0%
2	150	0	0.0%

3	198	0	0.0%
---	-----	---	------

**Distribución por Comuna (Top 10):**

Comuna	Total	Casos TEA	% TEA
Comuna 01	12	0	0.0%
Comuna 17	6	0	0.0%
Comuna 64	7	0	0.0%

**Distribución por Institución Educativa (Top 10):**

Institución Educativa	Total	Casos TEA	% TEA
Asodisvalle	24	0	0.0%
I.E. Politécnico Municipal de Cali	7	0	0.0%
I.E. Técnico Comercial José María Vivas Balcázar	7	0	0.0%

#### 4. Conclusiones y Recomendaciones

- Priorizar estrategias de apoyo y formación docente en función del grado, edad, comuna e IEO.
- Mejorar la calidad de los datos mediante la homologación de nombres de IEO y comunas, estandarización de grados, tratamiento de valores faltantes y definición de reglas claras para la marcación de TEA.
- Integrar los datos con la gestión pedagógica, incluyendo el PIAR, la participación familiar y el acompañamiento de IPS. Se recomienda además el desarrollo de tableros de control y series históricas para el seguimiento continuo.