



PROYECTO FINAL

ANALISIS EXPLORATORIO DE DATOS

**Población de Estudiantes con Autismo en Cali Registrados en el SIMAT-
Sistema de Matrícula Estudiantil de Educación Básica y Media**

MAESTRIA EN IA APLICADA

MATERIA: PROYECTO 1 DE INNOVACION TECNOLOGICA EN IA

PROFESOR: Ing. MILTON ORLANDO SARRIA PAJA

ALUMNOS:

Edwin Pérez

Cristian Quebrada

Rubén Darío Sabogal

Octubre de 2025 – Cali

Análisis Exploratorio de Datos para la población de niños Autistas escolarizados registrados en el SIMAT – Sistema de Matricula académica de educación preescolar, básica primaria y secundaria



Ministerio de
Educación Nacional
República de Colombia

Libertad y Orden



Sistema de Matricula Estudiantil de Educación Básica y Media

Usuario:

Contraseña:

ENTRAR

[Olvidé mi contraseña](#)



Versión 7.0.9.21 generada el 31/05/2025 03:27 PM SIMAT_FRONT_32

Análisis Exploratorio de Datos (EDA) del Dataset SIMAT Cali (Población TEA)

Objetivo: Realizar un análisis exploratorio de datos del dataset "POBLACIÓN TEA CALI.xlsx" para comprender las características clave de la población estudiantil identificada en el sistema SIMAT, con un enfoque particular en información relevante para la posible identificación temprana del Trastorno del Espectro Autista (TEA) en Cali.

Informe Ejecutivo

El análisis exploratorio de datos (EDA) de la población estudiantil en Cali identificada en el sistema SIMAT revela información demográfica, socioeconómica y educativa importante. La distribución de edades en el dataset muestra una concentración de estudiantes en rangos de edad escolar típica, aunque se identificaron algunos valores atípicos de mayor edad que podrían requerir una investigación adicional.

Database “ población_autista_cali”

ANO	INF	RUN CODIGO	CODIGO DANE	DANE ANTERIOR	NOMBRE INSTITUCION	COMUNA	CONS SEDE	NOMBRE SEDE	COMUNA SEDE	SI
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA JUANA DE CAicedo Y CUERO	COMUNA 20	1.76001E+13	INSTITUCION EDUCATIVA JUANA DE CAicedo Y CUERO	COMUNA 20	OF
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA ALFONSO LOPEZ PUMAREJO	COMUNA 07	1.76001E+13	INSTITUCION EDUCATIVA ALFONSO LOPEZ PUMAREJO	COMUNA 07	OF
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA TECNICO INDUSTRIAL COMUNA 17	COMUNA 17	1.76001E+13	SEDE LUIS CARLOS ROJAS GARCÉS	COMUNA 17	OF
2025	1	1.76002E+11	1.76002E+11	1.76002E+11	INSTITUCION EDUCATIVA LLANO VERDE	COMUNA 15	1.76002E+13	CALIMIO NORTE	COMUNA 06	NI
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA AGUSTIN NIETO CABALLERO	COMUNA 11	1.76001E+13	MARINO RENGIFO SALCEDO	COMUNA 11	OF
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA TECNICO INDUSTRIAL VEINTE DE JULIO	COMUNA 04	1.76001E+13	INSTITUCION EDUCATIVA TECNICO INDUSTRIAL VEINTE DE JULIO - SEDE PRINCIPAL	COMUNA 04	OF
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA LICEO DEPARTAMENTAL	COMUNA 19	1.76001E+13	INSTITUCION EDUCATIVA LICEO DEPARTAMENTAL - SEDE PRINCIPAL	COMUNA 19	OF
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA CARLOS HOLGUIN LLOREDA	COMUNA 10	1.76001E+13	INSTITUCION EDUCATIVA CARLOS HOLGUIN LLOREDA - SEDE PRINCIPAL	COMUNA 10	OF
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA TECNICO INDUSTRIAL VEINTE DE JULIO	COMUNA 04	1.76001E+13	INSTITUCION EDUCATIVA TECNICO INDUSTRIAL VEINTE DE JULIO - SEDE PRINCIPAL	COMUNA 04	OF
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA TECNICA COMERCIAL LAS AMERICAS	COMUNA 08	1.76001E+13	RAFAEL URIBE URIBE	COMUNA 08	OF
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA INEM JORGE ISAACS	COMUNA 04	1.76001E+13	INSTITUCION EDUCATIVA INEM JORGE ISAACS	COMUNA 04	OF
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA TECNICO COMERCIAL JOSE MARIA VIVAS BALCAZAR	COMUNA 10	1.76001E+13	INSTITUCION EDUCATIVA TECNICO COMERCIAL JOSE MARIA VIVAS BALCAZAR	COMUNA 10	OF
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA GENERAL FRANCISCO DE PAULA SANTANDER	COMUNA 11	1.76001E+13	JOSE VICENTE CONCHA	COMUNA 11	OF
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA REPUBLICA DE ISRAEL	COMUNA 04	1.76001E+13	INSTITUCION EDUCATIVA REPUBLICA DE ISRAEL	COMUNA 04	OF
2025	1	3.76001E+11	3.76001E+11	3.76001E+11	COLEGIO SANTA ISABEL DE HUNGRIA SEDE ALFONSO LOPEZ	COMUNA 07	3.76001E+13	COLEGIO SANTA ISABEL DE HUNGRIA SEDE ALFONSO LOPEZ	COMUNA 07	NI
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA LUIS FERNANDO CAICEDO	COMUNA 01	1.76001E+13	CECILIA CABALLERO DE LOPEZ	COMUNA 01	OF
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA DE SANTA LIBRADA	COMUNA 03	1.76001E+13	INSTITUCION EDUCATIVA DE SANTA LIBRADA	COMUNA 03	OF
2025	1	1.76002E+11	1.76002E+11	1.76002E+11	INSTITUCION EDUCATIVA GENERAL JOSE MARIA CABAL	COMUNA 18	1.76002E+13	INSTITUCION EDUCATIVA GENERAL JOSE MARIA CABAL SEDE PRINCIPAL	COMUNA 18	OF
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA TECNICO COMERCIAL JOSE MARIA VIVAS BALCAZAR	COMUNA 10	1.76001E+13	INSTITUCION EDUCATIVA TECNICO COMERCIAL JOSE MARIA VIVAS BALCAZAR	COMUNA 10	OF
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA NORMAL SUPERIOR LOS FARALLONES	COMUNA 03	1.76001E+13	INSTITUCION EDUCATIVA NORMAL SUPERIOR LOS FARALLONES	COMUNA 03	OF
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA AGUSTIN NIETO CABALLERO	COMUNA 11	1.76001E+13	MARINO RENGIFO SALCEDO	COMUNA 11	OF
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA REPUBLICA DE ISRAEL	COMUNA 04	1.76001E+13	INSTITUCION EDUCATIVA REPUBLICA DE ISRAEL	COMUNA 04	OF
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA TECNICO INDUSTRIAL JOSE MARIA CARBONELL	COMUNA 10	1.76001E+13	INSTITUCION EDUCATIVA TECNICO INDUSTRIAL JOSE MARIA CARBONELL	COMUNA 10	OF
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA JUAN DE AMPUDIA	COMUNA 08	1.76001E+13	INSTITUCION EDUCATIVA JUAN DE AMPUDIA	COMUNA 08	OF
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA SANTA FE	COMUNA 08	1.76001E+13	INSTITUCION EDUCATIVA SANTA FE	COMUNA 08	OF
2025	1	1.76002E+11	1.76002E+11	1.76002E+11	INSTITUCION EDUCATIVA LLANO VERDE	COMUNA 15	1.76002E+13	SAN LUIS	COMUNA 06	NI
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA MONSEÑOR RAMON ARCILA	COMUNA 14	1.76001E+13	INSTITUCION EDUCATIVA MONSEÑOR RAMON ARCILA - SEDE PRINCIPAL	COMUNA 14	OF
2025	1	3.76001E+11	3.76001E+11	3.76001E+11	liceos del EJERCITO PICHINCHA	COMUNA 18	3.76001E+13	liceos del EJERCITO PICHINCHA	COMUNA 18	SE
2025	1	1.76001E+11	2.76001E+11	2.76001E+11	INSTITUCION EDUCATIVA ISAIAS GAMBOA	COMUNA 01	1.76001E+13	EL AGUACATAL	COMUNA 01	OF
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA NORMAL SUPERIOR SANTIAGO DE CALI	COMUNA 10	1.76001E+13	INSTITUCION EDUCATIVA NORMAL SUPERIOR SANTIAGO DE CALI	COMUNA 10	OF
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA CELMIRA BUENO DE OREJUELA	COMUNA 05	1.76001E+13	INSTITUCION EDUCATIVA CELMIRA BUENO DE OREJUELA	COMUNA 05	OF
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA TECNICO INDUSTRIAL JOSE MARIA CARBONELL	COMUNA 10	1.76001E+13	INSTITUCION EDUCATIVA TECNICO INDUSTRIAL JOSE MARIA CARBONELL	COMUNA 10	OF
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA CARLOS HOLMES TRUJILLO	COMUNA 16	1.76001E+13	UZANDRO FRANKY	COMUNA 16	OF
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA TECNICO DE COMERCIO SANTA CECILIA	COMUNA 02	1.76001E+13	INSTITUCION EDUCATIVA TECNICO DE COMERCIO SANTA CECILIA - SEDE PRINCIPAL	COMUNA 02	OF
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA EVA RIASCOS PLATA	COMUNA 12	1.76001E+13	INSTITUCION EDUCATIVA EVA RIASCOS PLATA - SEDE PRINCIPAL	COMUNA 12	OF
2025	1	1.76001E+11	1.76001E+11	1.76001E+11	INSTITUCION EDUCATIVA NORMAL SUPERIOR SANTIAGO DE CALI	COMUNA 10	1.76001E+13	INSTITUCION EDUCATIVA NORMAL SUPERIOR SANTIAGO DE CALI	COMUNA 10	OF

Diccionario de Datos:

DICCIONARIO DE DATOS - POBLACION AUTISMO - SIMAT		
VARIABLE	VALOR	SIGNIFICADO
ANO_INF	Numerica	Año Escolar
MUN_CODIGO	Numerica	Codigo de Municipio
CODIGO_DANE	Numerica	Codigo DANE
DANE_ANTERIOR	Numerica	DANE anterior
NOMBRE_INSTITUCION	Categorica	Nombre de la Institucion
COMUNA	Numerica	Comuna
CONS_SEDE	Numerica	Consecutivo Sede
NOMBRE_SEDE	Categorica	Nombre de la Sede
COMUNA_SEDE	Numerica	Comuna de la Sede
SECTOR	Categorica	Sector /Barrio
ZONA	Numerica	Zona
TIPO_DOCUMENTO	Categorica	Tipo de Documento
NRO_DOCUMENTO	Numerica	Numero de Documento
EXP_DEPTO	Categorica	Departamento de Expedicion
EXP_MUN	Categorica	Numero de Expedicion
APELLIDO1	Categorica	Primer Apellido
APELLIDO2	Categorica	Segundo Apellido
NOMBRE1	Categorica	Primer Nombre
NOMBRE2	Categorica	Segundo Nombre
DIRECCION_RESIDENCIA	Categorica	Direccion de Residencia
TEL	Numerica	Telefono
RES_DEPTO	Categorica	Departamento de Residencia
RES_MUN	Categorica	Municipio de Residencia
ESTRATO	Numerica	Estrato
SISBEN_IV	Numerica	Numero de SISBEN
FECHA_NACIMIENTO	Fecha	Fecha de Nacimiento
NAC_DEPTO	Categorica	Departamento de Nacimiento
NAC_MUN	Categorica	Municipio de Nacimiento
GENERO	Categorica	Genero
POB_VICT_CONF_RUV	si/no	Poblacion Victima del Conflicto
PROVIENE_SECTOR_PRIV	si/no	Proviene de Institucion Privada
PROVIENE_OTR_MUN	si/no	Proviene de Otro Municipio
TIPO_DISCAPACIDAD	Categorica	Tipo de Discapacidad
CAP_EXC	Categorica	Capacidad Extraordinaria
ETNIA	Categorica	Grupo Etnico
RES	Numerica	
INS_FAMILIAR	Categorica	Institucion Familiar
TIPO_JORNADA	Categorica	Tipo de Jornada
CARACTER	Numerica	Carácter
ESPECIALIDAD	Categorica	Especialidad
GRADO	Numerica	Grado Escolar
GRUPO	Categorica	Grupo Escolar
METODOLOGIA	Categorica	Metodologia
MATRICULA_CONTRATADA	Categorica	Matricula
REPITENTE	Categorica	Repitente
NUEVO	Categorica	Estudiante Nuevo
FUE_RECU	Numerica	
ZON_ALU	Categorica	Zona Escolar Alumno
CAB_FAMILIA	Categorica	
BEN_MAD_FLIA	Categorica	
BEN_VET_FP	Categorica	
BEN_HER_NAC	Categorica	
CODIGO_INTERNADO	Numerica	Codigo Internado
CODIGO_VALORACION_1	Numerica	Codigo de valoracion 1
CODIGO_VALORACION_2	Numerica	Codigo de Valoracion 2
NUM_CONVENIO	Numerica	Numero de Convenio
PER_ID	Numerica	
APOYO_ACADEMICO_ESPECIAL	Categorica	Apoyo Academico Especial
SRPA	Numerica	
PAIS_ORIGEN	Categorica	Pais de origen
TRASTORNOS_ESPECIFICOS	Categorica	Trastornos Especificos

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 520 entries, 0 to 519
Data columns (total 61 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   ANO_INF                520 non-null    int64  
1   MUN_CODIGO             520 non-null    int64  
2   CODIGO_DANE             520 non-null    int64  
3   DANE_ANTERIOR          520 non-null    int64  
4   NOMBRE_INSTITUCION     520 non-null    object  
5   COMUNA                 520 non-null    object  
6   CONS_SEDE              520 non-null    int64  
7   NOMBRE_SEDE            520 non-null    object  
8   COMUNA_SEDE            520 non-null    object  
9   SECTOR                 520 non-null    object  
10  ZONA                   520 non-null    object  
11  TIPO_DOCUMENTO         520 non-null    int64  
12  NRO_DOCUMENTO          520 non-null    object  
13  EXP_DEPTO              520 non-null    int64  
14  EXP_MUN                520 non-null    int64  
15  APELLIDO1              520 non-null    object  
16  APELLIDO2              520 non-null    object  
17  NOMBRE1                 520 non-null    object  
18  NOMBRE2                 322 non-null    object  
19  DIRECCION_RESIDENCIA   520 non-null    object  
...
59  PAIS_ORIGEN            520 non-null    int64  
60  TRASTORNOS_ESPECIFICOS 520 non-null    int64  
dtypes: datetime64[ns](1), float64(3), int64(30), object(27)
memory usage: 247.9+ KB

Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...
```

Análisis Bivariado y Clustering de la Población Estudiantil con TEA en Cali

El análisis bivariado entre las variables estrato socioeconómico y edad revela que, aunque la mayoría de los estudiantes pertenecen a estratos bajos y medios, no se identifica una correlación lineal significativa entre ambas variables. No obstante, se observa que los casos con edades atípicas tienden a concentrarse en los estratos más bajos, lo cual sugiere la necesidad de investigar el contexto particular de estos estudiantes.

Por otro lado, mediante la técnica de clustering K-Means, utilizando las variables Edad, Estrato, Comuna y Grado, se logró segmentar la población en tres grupos diferenciados. Estos clusters presentan perfiles distintos en cuanto a edad, nivel socioeconómico y grado escolar, lo que evidencia la diversidad dentro de la población registrada en el sistema SIMAT. Además, la distribución geográfica por comuna dentro de cada grupo permite identificar zonas donde ciertos perfiles son más frecuentes.

Principales Hallazgos

Edad (Análisis Univariado): La edad promedio de los estudiantes es de 10.97 años, con una mediana cercana a los 10 años y una moda de 8 años. Se detectaron dos valores atípicos con edades de 31 y 27 años. La distribución presenta varios picos, lo que podría indicar la existencia de subgrupos dentro de la muestra.

Estrato vs. Edad (Análisis Bivariado): No se evidencia una relación lineal clara entre estas variables. Los estudiantes con edades atípicas pertenecen a estratos bajos, lo que podría reflejar trayectorias educativas no convencionales o participación en programas especiales. La mayoría se concentra en los estratos 1, 2 y 3.

Segmentación por Clustering (K-Means):

Cluster 0: Estudiantes de mayor edad (promedio de 15 años), en grados avanzados (promedio de grado 9), distribuidos en diversas comunas y pertenecientes a estratos bajos y medios (promedio de estrato 2).

Cluster 1: Niños más jóvenes (promedio de 8.5 años), en grados iniciales (promedio de grado 2), concentrados en estratos bajos (promedio de estrato 1.5) y distribuidos en varias comunas.

Cluster 2: También conformado por estudiantes jóvenes (promedio de 8.8 años) en grados iniciales (promedio de grado 2.3), pero con un estrato promedio más alto (3.3) y una distribución geográfica distinta a la del Cluster 1.

Implicaciones para la Política Pública

La segmentación obtenida mediante clustering ofrece una base sólida para orientar decisiones estratégicas en el ámbito de la política pública, especialmente en relación con el Trastorno del Espectro Autista (TEA):

Detección Temprana: Los Clusters 1 y 2, compuestos por estudiantes en etapas escolares iniciales, representan una población clave para implementar programas de detección temprana y sensibilización en las comunas donde estos perfiles predominan.

Intervenciones Adaptadas: Los distintos perfiles sugieren la necesidad de diseñar estrategias diferenciadas. El Cluster 0 podría beneficiarse de programas de apoyo académico y social para adolescentes, mientras que los Clusters 1 y 2 requieren intervenciones centradas en el desarrollo de habilidades comunicativas y sociales desde edades tempranas, en línea con los objetivos del proyecto SMART.

Asignación Eficiente de Recursos: Identificar comunas con alta concentración de estudiantes en los clusters de mayor riesgo (especialmente Clusters 1 y 2) permite una distribución más estratégica de recursos, como profesionales especializados (psicólogos, terapeutas, fonoaudiólogos), materiales educativos adaptados y capacitación docente.

Análisis de Casos Atípicos: La presencia de estudiantes mayores en estratos bajos dentro del Cluster 0 requiere una investigación individualizada para comprender las posibles barreras que han afectado su trayectoria educativa y diseñar medidas preventivas.

Promoción de la Equidad: Las diferencias en el estrato socioeconómico entre los clusters jóvenes sugieren que el contexto puede influir en el acceso a servicios y diagnóstico. Es

fundamental garantizar igualdad de oportunidades para todos los estudiantes, independientemente de su nivel socioeconómico.

Recomendaciones para Futuros Pasos

Ampliar el análisis incorporando variables adicionales del dataset SIMAT (tipo de discapacidad, apoyos académicos, etc.).

Integrar datos externos (salud, rendimiento académico) para enriquecer los perfiles de los clusters.

Validar los resultados del clustering con expertos y actores locales.

Utilizar los clusters como insumo para desarrollar modelos predictivos de riesgo de TEA.

Diseñar planes de acción concretos y medibles para implementar las recomendaciones de política pública derivadas del análisis.