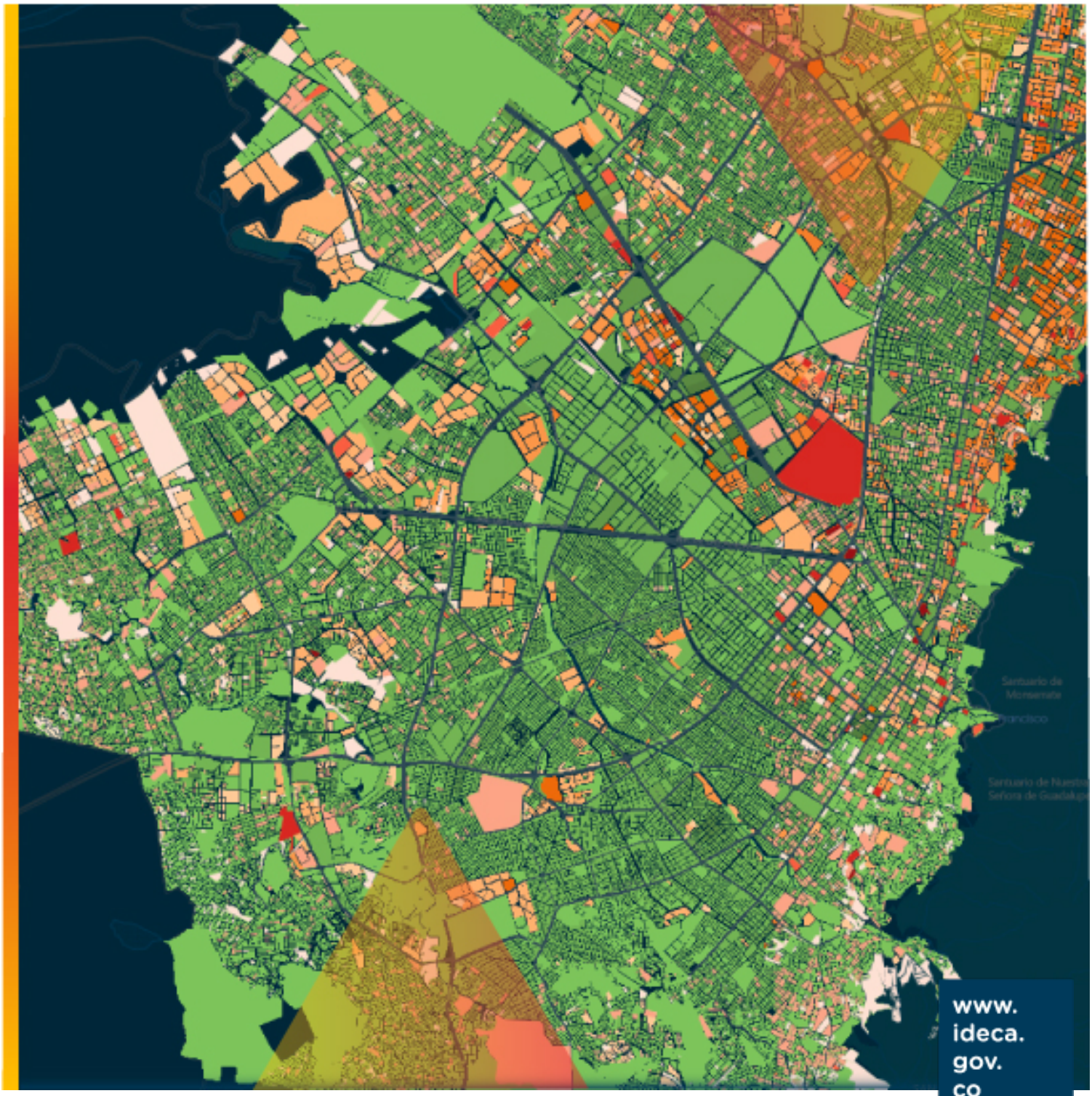
ideca



Implementación del programa o plan de incubación
-Estrategia pedagógica de artificación en el uso reflexivo de los datos-

—

Bogotá, D.C., 31/12/2025

ideca

Implementación del programa o plan de Incubación -Estrategia pedagógica de artificación en el uso reflexivo de los datos-



Gerente IDECA
Salomé Naranjo Luján

Subgerente de Analítica de Datos
Diego Ricardo Ibarra Rodríguez

Profesional Especializado
Martha Nubia Rojas Guerrero

Profesional Universitario
John Fredy Monje Mahete

Profesional Analista de Datos
Lilian Daniela Beltrán Barrera

Fecha de creación o actualización: 31/12/2025

Página web: www.ideca.gov.co

Correo electrónico: ideca@catastrobogota.gov.co

Licencia: Attribution 4.0 International (CC BY 4.0)

ALCALDÍA MAYOR DE BOGOTÁ D.C.

ideca

Control de Versiones

CAMBIOS

Fecha	Autor	Versión	Cambio efectuado
30/09/2025	Martha Nubia Rojas G.	1.0	Primera versión del documento. Presentación Tabla de Contenido. No hay cambios para registrar. **(Documento en construcción)**
30/10/2025	Martha Nubia Rojas G.	2.0	Se incluye ficha descriptiva del Proyecto Biomasa.
30/11/2025	Martha Nubia Rojas G.	3.0	Se incluye i) Análisis descriptivos univariados a los datos suministrados; y ii) Procesamiento de datos.
31/12/2025	Martha Nubia Rojas G.	4.0	Se incluye ficha descriptiva del Proyecto Paneles Solares.

REVISORES

Nombre	Dependencia
Diego Ricardo Ibarra R.	Subgerencia de Analítica de Datos

Tabla de Contenido

Resumen ejecutivo **6**

Alcance **7**

Objetivos **7**

1. Plan o Programa de Incubación **8**
1.1. Identificación de Iniciativas 8
1.2. Elaboración de Fichas Descriptivas 9
1.3. Análisis de alternativas de solución de la iniciativa 38
1.4. Concretar la solución 38
1.5. Avalar la idea 39
1.6. Identificación de Equipos para desarrollar la iniciativa 40
1.7. Solucionar la Iniciativa 41

2. Conclusiones **70**

3. Anexo 1: Tipologías Edad Escolar **71**

Lista de tablas

Tabla 1. Identificación de Iniciativas potenciales a ser Incubadas o Aceleradas 8
Tabla 2. Ficha Descriptiva del Proyecto – **SED - Obce** 9
Tabla 3. Ficha Descriptiva del Proyecto Electrolineras – **Universidad Nacional de Colombia** 19
Tabla 4. Ficha Descriptiva del Proyecto Biomasa – **Universidad Nacional de Colombia** 23
Tabla 5. Ficha Descriptiva del Proyecto Paneles Solares – **Universidad Nacional de Colombia** 30
Tabla 6. Equipo de Trabajo para desarrollar el proyecto de incubación 40
Tabla 7. Variables Explicativas y Variables Respuesta 42
Tabla 8. Variables Explicativas y Variables Respuesta 47

Lista de gráficas

Gráfica 1. Análisis inicial de las variables 42
Gráfica 2. Variables explicativas 43
Gráfica 3. Análisis inicial de las variables (por jornada, localidad, tipo colegio, grado) 44
Gráfica 4. Análisis inicial – Variable explicativa Edad 45
Gráfica 5. Análisis Variables Respuesta 46
Gráfica 6. Análisis del tipo de violencia respecto a la jornada 47
Gráfica 7. Análisis Tipo de Violencia respecto a los Rangos de Edad 48
Gráfica 8. Análisis Tipo de Violencia respecto al Nivel Académico 49
Gráfica 9. Análisis Tipo de Violencia respecto a la Localidad 50

Gráfica 10. Análisis Tipo de Violencia respecto al Género 51
Gráfica 11. Variables del Modelo (Explicativas y Respuesta) 52
Gráfica 12. Diseño de Prototipo Entregable 52
Gráfica 13. Diseño de Prototipo Entregable 53
Gráfica 14. Experimentos realizados 55
Gráfica 15. Datos de las violencias desbalanceados 55
Gráfica 16. Mejor modelo "XGBOOST" 56
Gráfica 17. Ajuste de hiperparámetros al mejor modelo XGBOOST 57
Gráfica 18. Sobremustreo "Técnica Smote" 58
Gráfica 19. Submuestreo 59
Gráfica 20. XGBOOST con el Modelo Base 59
Gráfica 21. XGBOOST con el Modelo Over-Under 60
Gráfica 22. XGBOOST con mejor Accuracy y f1-score 61
Gráfica 23. Tablero de predicción para Abuso y Violencia 62
Gráfica 24. Tablero de predicción para Abuso y Violencia – Módulo carga de datos 63
Gráfica 25. Módulo carga de datos – Ejemplo datos cargados 63
Gráfica 26. Tablero de predicción para Abuso y Violencia – Módulo análisis descriptivo 64
Gráfica 27. Tablero de predicción para Abuso y Violencia – Distribución de variables categóricas 64
Gráfica 28. Tablero de predicción para Abuso y Violencia – Distribución por Localidad 65
Gráfica 29. Tablero de predicción para Abuso y Violencia – Tabla de reportes 65
Gráfica 30. Tablero de predicción para Abuso y Violencia – Módulo resultados de predicción 66
Gráfica 31. Distribución de la predicción tipo violencia 66
Gráfica 32. Distribución de la predicción tipo agresor 67
Gráfica 33. Tablero de predicción para Abuso y Violencia – Despliegue en Azure - 67
Gráfica 34. Módulos Tablero de predicción para Abuso y Violencia – Despliegue en Azure - 68

Resumen ejecutivo

En el marco del Plan Anual de Trabajo IDECA 2025, Objetivo Estratégico 3 (OE3): Generar valor a los usuarios con innovación, creatividad y uso de la tecnología, se ha formulado como Línea de Acción 3.1 (LA 3.1): Establecer un programa de incubación de iniciativas que se enfoquen en el desarrollo de soluciones de valor basadas en la información geoespacial, con la **Meta/Producto "Dos (2) ejercicios de incubación y aceleración de soluciones geoespaciales".**

Por lo anterior, este documento **consolida el desarrollo de uno de los ejercicios de incubación con la aplicación de cada una de las etapas** establecidas para el **Proyecto de Incubación "Estrategia pedagógica de artificación en el uso reflexivo de los datos",** priorizado de acuerdo con los resultados de la Matriz de Priorización.

Las etapas desarrolladas son:

i) Identificación de iniciativas.
ii) Elaboración de Fichas Descriptivas
iii) Análisis de alternativas de solución de la iniciativa.
iv) Concretar la solución.
v) Avalar la idea.
vi) Identificación de Equipos para desarrollar la iniciativa.
vii) Solucionar la iniciativa.

Alcance

El Proyecto de Incubación "Estrategia pedagógica de Artificación en el uso reflexivo de los datos" se desarrollará en 7 meses como un proyecto "piloto", donde se analizará la data suministrada por la Secretaría de Educación a través del Observatorio de Convivencia Escolar; luego, se desarrolla un procesamiento de datos; y como resultado final se propone una solución de analítica desarrollada a través de un modelo predictivo multietiqueta implementado a través de un aplicativo desarrollado en Python.

Objetivos

Objetivo General

Desarrollar un modelo predictivo multietiqueta que, a partir de variables sociodemográficas (como edad, género, localidad y UPZ), permita predecir las características del posible suceso de abuso o violencia que podría experimentar un estudiante.

Objetivos Específicos

Aplicar técnicas de limpieza y estructuración de datos durante el procesamiento de los mismos.

Utilizar como variables respuesta aspectos como el lugar de ocurrencia, el tipo de violencia, la relación con el agresor y los hechos violentos.

Establecer asociaciones que permitan identificar perfiles de riesgo y apoyar la toma de decisiones en prevención y atención.

Realizar mesas de trabajo técnicas con la Secretaría de Educación Distrital y con el Observatorio de Convivencia Escolar.

1. Plan o Programa de Incubación

Esta sección describe la ejecución de cada una de las etapas establecidas en el Plan de Incubación y Aceleración de Soluciones Geoespaciales así:

1.1. Identificación de Iniciativas

El proceso de indagación de iniciativas susceptibles de ser incubadas o aceleradas se realizó mediante la aplicación de la encuesta titulada **"Recepción de retos para evaluar la aplicación de la Analítica de Datos"** enviada a las organizaciones que se tabulan a continuación:

T1 Tabla 1. Identificación de Iniciativas potenciales a ser Incubadas o Aceleradas
Fuente: Elaboración propia.

No.	Entidad donde desempeña sus funciones	Dependencia /área	Nombres y apellidos	Correo electrónico
1	Secretaría de Educación (SED)	Observatorio de Convivencia Escolar (Obce)	Edwin Alberto Ussa Cristiano	eussa@educacionbogota.gov.co
2	Universidad Nacional de Colombia	Facultad de Ingeniería (Ingeniería Eléctrica)	Javier Alveiro Rosero García	jaroserog@unal.edu.co
3	Universidad Nacional de Colombia	Facultad de Ingeniería (Ingeniería Eléctrica)	Oscar Perdomo	ojperdomoc@unal.edu.co
4	Universidad Nacional de Colombia	Facultad de Ingeniería (Ingeniería Eléctrica)	Javier Alveiro Rosero García	jaroserog@unal.edu.co
5	IDRD	Subdirección Técnica de Construcciones	Julián Aristizábal O.	julian.aristizabal@idrd.gov.co

1.2. Elaboración de Fichas Descriptivas

Una vez recibidas las encuestas con las posibles iniciativas a ser incubadas o aceleradas, se procede a elaborar las respectivas fichas descriptivas, así:

T2 Tabla 2. Ficha Descriptiva del Proyecto – **SED - Obce**
Fuente: Elaboración propia.

Ficha Descriptiva – Identificación Caso de Uso – Subgerencias de Analítica de Datos			
Información de la Entidad			
Entidad/ Organización	Secretaría de Educación Distrital		
Dependencia/área	Oficina para la Convivencia Escolar		
Enlace	**Correo electrónico**	**Teléfono**	**Cargo**
Angela Viviana Roa Ruíz	avroa@educacionbogota.gov.co	n/a	Contratista
Información del reto			
Defina cuál es el reto (necesidad, problemática o deseo). Escriba su reto usando la siguiente estructura gramatical: cómo + verbo en infinitivo + complemento. Ejemplo: cómo identificar permanentemente las zonas con mayores casos o situaciones de violencia infantil en Bogotá.	¿Cómo diseñar e implementar una estrategia pedagógica de artificación para el uso reflexivo de datos en la generación de conocimientos sobre convivencia escolar con y desde las comunidades educativas en Bogotá? En específico, con el análisis de las particularidades territoriales asociadas a la prevalencia de violencias hacia NNJ. Comprendemos el territorio como un espacio físico y social, el cual abarca no solo delimitaciones geográficas, sino también relaciones sociales y de apropiación de dicho espacio. De ahí que partamos de la hipótesis de que las características de estos espacios podrían influir en la prevalencia o no de violencias contra niñas, niños y adolescente. Si bien desde la SED hemos avanzado en la caracterización de los entornos escolares a través de la identificación de factores de riesgo como mala		

	disposición de basuras, venta y consumo de sustancias psicoactivas, acoso sexual callejero, riñas, ausencia de señalización escolar, infraestructura y cuerpos de agua, puntos de atraco y presencia de habitabilidad en calle; vemos pertinente fortalecer el análisis de la relación entre estas características espaciales y la prevalencia de violencias reportadas en el Sistema de Alertas. Asimismo, sería positivo revisar la posible relación existente entre otras particularidades territoriales como las sociodemográficas (Índice pobreza, concentración del desempleo; actividades económicas), relacionales (Violencia en el contexto familiar; Violencias Basadas en Género; ausencia y presencia de equipamientos urbanos; características ambientales (ej: humedales, curtiembres), con la prevalencia de violencias reportadas en el sistema de alertas. Es importante tener en cuenta que el módulo de abuso y violencias del Sistema de Alerta incluye el reporte de seis violencias: física, sexual, psicológica, negligencia, económica y abandono. No obstante, para este estudio, concentraremos en las cuatro primeras dado el aumento en el reporte durante 2024.
Descripción detallada del reto. Cuéntenos más sobre éste, el contexto del mismo, indicadores actuales si los hay, cómo lo resuelve actualmente. Si es un deseo explique que le gustaría tener, cómo se imagina la solución ideal, otros.	En el año 2024 el Observatorio creó Bogodatos para la Convivencia Escolar en sus versiones trimestrales, semestrales y anuales, un material que analiza las principales cifras del Sistema de Alertas sobre seis módulos: abuso y violencia, conducta suicida, consumo de sustancias psicoactivas, maternidad y paternidad temprana, accidentalidad, trastornos del aprendizaje y el comportamiento, y entornos escolares. Si bien este ejercicio ha incluido un análisis descriptivo de datos, la experiencia evidencia la necesidad de desarrollar un análisis estadístico “avanzado” que permita un mejor aprovechamiento de la data, para la interpretación, estudios y elaboración de recomendaciones de política pública. La Secretaría de Educación tiene un interés en la compresión de los entornos educativos lo cual implica una aproximación a las particularidades territoriales y su incidencia en la prevalencia de violencias hacia NNJ. A partir de ello, el OBCE en el material Bogodatos para la Convivencia escolar, ha avanzado en un análisis

	territorial de las situaciones que afectan la convivencia escolar (módulos del sistema de alertas), que incluye el cálculo de la tasa por cada 1.000 estudiantes en cada una de las localidades de Bogotá, y la elaboración de unos mapas de calor que visualizan los hallazgos. No obstante, este análisis se ha desarrollado exclusivamente con los reportes del sistema de alertas y no ha incluido variables geográficas diversas, ni tampoco se ha desarrollado análisis estadísticos predictivos y/o correlacionales con características espaciales de la ciudad. El conocimiento de las particularidades territoriales y su incidencia en la generación de violencias podría tener un efecto positivo en la aproximación de las comunidades educativas a las problemáticas de convivencia escolar. Esto, debido a que proporciona datos situados que pueden impactar no solo en su visibilización, sino en la creación de espacios de reflexión pedagógica y construcción de conocimiento para la transformación desde dichas comunidades. A esto se suma, los necesidades y retos identificados en los ejercicios de difusión y apropiación social del material con las comunidades educativas: (i) uso de estrategias pedagógicas diferenciadas por cursos de vida y para diversos actores de la comunidad educativa, en el abordaje de los datos; (ii) fomento de la construcción de conocimiento que contribuya a la prevención de violencias y situaciones que afectan la salud mental y el bienestar de las y los estudiantes. Lo anterior plantea desafíos concernientes a quién, cómo y para qué generar conocimiento sobre datos de convivencia escolar, los cuales se pretenden resolver en esta investigación. Frente a ello, el OBCE ha avanzado en ejercicios de “datos a la calle” que consisten en la socialización y difusión del material con docentes, estudiantes, y madres, padres, cuidadores, etc. Esto a partir del uso de Bogodatos para la Convivencia Escolar y metodologías diversas como la “corografía emocional”, la cual ha posibilitado compartir emociones que surgen en los participantes, luego de aproximarse a los reportes de conducta suicida. Asimismo, se han desarrollado ejercicios de percepción y contraste de

	datos con la información publicada en los materiales del OBCE, que ha fomentado el (auto)análisis de los imaginarios de los participantes, y ha permitido la apropiación de los hallazgos de los boletines dada la naturaleza de las metodologías y dinamizado la reflexión sobre la corresponsabilidad en la garantía de derechos. Si bien hemos avanzado en el desarrollo de los desafíos mencionados, consideramos que aún existe un amplio camino en la pedagogía del dato, pues existen pocos referentes en Bogotá que además de la implementación de metodologías innovadoras, tenga en cuenta aspectos éticos y principios de cuidado con las comunidades. Por eso, desde el OBCE esperamos desarrollar "laboratorios de artificación del dato" como una apuesta metodológica que garantice la participación
¿Por qué es importante para usted este reto y para qué necesita o quiere resolverlo?	El desarrollo del reto es necesario para avanzar en los procesos de humanización del dato, los cuales requieren de respuestas axiológicas, epistemológicas y ontológicas. Las primeras hacen referencia al abordaje de valores, principios éticos y de cuidado para el análisis, la difusión y apropiación de la información, ya que desde el OBCE trabajamos con datos sensibles (violencias, conducta suicida, maternidades tempranas, consumo de sustancias psicoactivas). Las respuestas epistemológicas implican profundizar en qué tipo de conocimiento estamos produciendo, y cómo estamos comprendiendo esto que llamamos convivencia escolar a través de los datos. Las respuestas ontológicas invitan a cuestionar quién o quiénes producen el conocimiento y para qué, por ello, para el OBCE la inclusión de las comunidades educativas, en particular de las voces de los NNJ en los procesos de análisis e investigación de los datos, es fundamental. Así, el diseño e implementación de estrategias de artificación para la reflexión pedagógica de los datos, requiere de cuestionar la naturaleza de nuestro quehacer en el OBCE, pero, sobre todo, comprender que los datos con los que trabajamos abarcan vidas y afectaciones tanto a los NNJ como a sus

	familias, comunidades barriales y colegios. En otras palabras, la pedagogía del dato permite ir más allá de la cifra.
¿Considera que la solución a su reto puede apoyarse con analítica de datos? Responda sí, no, depende o supongo y explique su respuesta.	Tal como se menciona en la respuesta a la segunda pregunta, sí requerimos un proceso de analítica de datos que permita la articulación entre la información que tenemos en el Sistema de Alertas sobre violencias, y otras fuentes de información espaciales de la ciudad. Esto con el fin de identificar las particularidades territoriales que pueden incidir en una mayor prevalencia de violencias hacia NNJ, y que sirva como insumo fundamental en el desarrollo de la estrategia pedagógica de artificación. El proceso de analítica de datos también es importante para el OBCE en su interés por mejorar sus herramientas de análisis, interpretación y visualización de datos. Para este último, encontramos relevante explorar las tendencias de Data-art y artificación del dato, que, si bien son recientes, plantean unas respuestas innovadoras a la democratización y comprensión de los datos abiertos.
¿Qué tipo de analítica de datos considera se debe desarrollar en la solución de su reto?	Combinación de varias (diagnóstica y predictiva).
¿La solución, resultado o entregable tendría algún impacto para la ciudad? Justifique su respuesta.	Análisis de fuentes de información interinstitucionales En la actualidad, se están desarrollando ejercicios de construcción de “lago de datos” y avanzando en la inter-operatividad de fuentes de información del distrito operadas por entidades diversas, para posibilitar la investigación integral y favorecer la garantía de derechos de NNJ. El ejercicio de analítica de datos, además de ser insumo para la construcción de los laboratorios de artificación, podría ser una experiencia exitosa, que fortalezcan la importancia de realizar análisis que incluyan más de una fuente de información, en correspondencia con las experiencias que vienen avanzando en la ciudad. Asimismo, la identificación de patrones territoriales que posibilitan la prevalencia de violencias hacia los NNJ, permitiría elaborar recomendaciones de política pública que incluyan a varias entidades del distrito. Desarrollo de metodologías, sistematización de la experiencia y protocolo de uso ético de datos Además, este proyecto

	contempla la elaboración de un material con metodologías para el abordaje pedagógico y artístico del dato, el cuál será insumo para los ejercicios de apropiación social del conocimiento sobre convivencia escolar. Asimismo, se espera elaborar un documento con la sistematización de la experiencia que incluirá el desarrollo de un protocolo de uso ético de datos.
¿A qué objetivo estratégico, programa o artículo de su entidad u organización y/o del Plan Distrital de Desarrollo está relacionado su reto?	A nivel distrital, esta investigación se enmarca en los ejes “Bogotá confía en su potencial” y “Bogotá confía en su gobierno” del Plan Distrital de Desarrollo (Secretaría Distrital de Planeación, 2024). Esto en la medida en que los datos sobre convivencia escolar en Bogotá y el conocimiento creado a partir de ellos pueden ser útiles para la prevenir y atender situaciones que afectan la integridad de las y los estudiantes, tal como lo establece el programa de atención integral a la primera infancia y educación como eje del potencial humano. Adicionalmente, la reflexión pedagógica de los datos sobre estas temáticas implica la existencia de un programa de gobierno abierto, íntegro, transparente y corresponsable, en el cual participa el Obce como miembro de la Red de Observatorios Distritales. Lo anterior se articula, a su vez, con la finalidad del Observatorio al ser este un espacio para “la investigación, la información, el análisis y la reflexión sobre temas relacionados a la convivencia, protección y seguridad escolar del Distrito escolar” (Acuerdo 434 de 2010, s.p.). Así, el abordaje pedagógico del dato para propiciar la construcción de conocimiento con y desde los miembros de las comunidades educativas acarrea el desarrollo de programas de investigación que den cuenta de formas para transformar las realidades educativas en los entornos educativos, familiares e individuales.
¿Quién(es) serán los beneficiarios de la solución? Puede haber varios usuarios y beneficiarios. Descríbalos brevemente.	Comunidades educativas (estudiantes y padres, madres, cuidadores)
¿Cuáles son las causas del reto (necesidad, problemáticas o deseos)?	1. Necesidad de desarrollar un análisis estadístico “avanzado” sobre las violencias ocurridas hacia NNJ, que permita un mejor aprovechamiento de la data para

	la interpretación, estudios y elaboración de recomendaciones de política pública sobre convivencia escolar. 2. Desarrollo de estrategias pedagógicas diferenciadas por cursos de vida y para diversos actores de la comunidad educativa, en el abordaje de los datos. 3. Fomento de la construcción de conocimiento situado que contribuya a la prevención de violencias y situaciones que afectan la salud mental y el bienestar de las y los estudiantes. 4. Desarrollo de un protocolo de uso ético de datos sobre convivencia escolar.
¿Cuáles son las consecuencias del reto?	1. Ausencia de una comprensión integral de las situaciones geográficas, económicas y estructurales relacionadas con la prevalencia de las violencias hacia NNJ en entornos escolares. 2. Ausencia de estrategias para el abordaje pedagógico del dato que tengan en cuenta las barreras en el conocimiento estadístico, la visibilización de situaciones que afectan los derechos de los NNJ, y la distancia en el reconocimiento del dato como un proceso humano que permite la toma de decisiones cotidianas. 3. Dificultades en la apropiación social del conocimiento por parte de las comunidades educativas alrededor de las investigaciones desarrolladas sobre situaciones que atienden y/o viven en su cotidianidad, haciendo que no se identifiquen sus potenciales usos. 4. Ausencia de principios y procesos éticos y cuidadosos con los datos sobre convivencia escolar, por parte de medios de comunicación, y diversos actores de la comunidad educativa.
¿Qué información sugiere o puede compartir para profundizar en el entendimiento del reto?	Recursos digitales, audiovisuales, web
¿Cuéntenos más sobre los detalles de conceptos, temas, documentos, fuentes de información, instructivos o bibliografía que nos sugiere o puede remitirnos al correo electrónico jmonje@catastrobogota.gov.co,	Desde el OBCE hemos avanzado en la construcción de un anteproyecto de investigación que establece el problema y pregunta de investigación, objetivos y justificación. Además, incluye una revisión de antecedentes con dos cuerpos de literatura, identifica

para profundizar en el entendimiento de su reto o expectativas de solución?	las categorías analíticas emergentes del proyecto, y un primer acercamiento metodológico.
Capacidades actuales	
¿Cuál es el equipo de trabajo del proyecto que trabaja o trabajará conjuntamente en resolver el reto. Especifique para cada uno: 1. Rol principal, 2. Profesión 3. Si es interno o externo, 4. Si tiene conocimiento o no, en analítica de datos.	Angela Roa - Investigadora - Antropóloga y licenciada en lenguas modernas - SED – OBCE - No analítica de datos Dina Riaño - Investigadora - Socióloga- SED – OBCE - Sí tiene conocimiento de analítica de datos Marcela Jiménez - Investigadora - Socióloga - SED – OBCE - No analítica de datos Stephania García - Líder de procesos de artificación - Artista - SED – OBCE - No analítica de datos
¿Qué datos o información dispone actualmente en su organización que sean importantes y se puedan usar para dar respuesta al reto identificado para generar la solución que se defina? Descríbalos brevemente: 1. Tipo de dato (estructurado, semiestructurado, no estructurado). 2. Tiene restricción o no para su acceso. 3. El origen y 4. Otros.	Disponemos de los datos estructurados del módulo de abuso y violencia del Sistema de Alertas de la Secretaría de Educación, que corresponde a casos de posibles vulneraciones identificados por docentes orientadores en los colegios oficiales y privados de Bogotá. Son accesibles (para visualización y descarga) por medio de la página del OBCE en el visor del Power Bi. Es importante mencionar, que las variables que comprometen datos sensibles no están abiertos, y para un potencial uso de estos, se deberá justificar éticamente la razón de su necesidad, ante la Oficina para la Convivencia Escolar. Los datos que no están públicos son los siguientes: Nombre del estudiante; ID del estudiante; Lugar de vivienda del estudiante; Información de contacto de los acudientes; Colegio; Documento de identificación; Descripción de los casos.
¿Dispone de recursos tecnológicos (Hardware o Software) para usarse durante el proceso de generación de la solución? Descríbalos brevemente.	Disponemos de los programas de Microsoft Office y una plataforma de Power Bi que permite la visualización de datos sobre convivencia escolar en Bogotá.
Actualmente en qué estado de avance o implementación tiene su idea o solución internamente en su entidad	Tengo un concepto por escrito y pregunte si es viable técnicamente, si funcionaría y si la aprueba la organización.

La solución imaginada	
¿Existe alguna solución o prototipo para resolver este reto que usted conozca o sea un referente?	Existen varios referentes con respecto a la relación entre arte dato y abordaje de datos sensibles. El primero abarca experiencias pedagógicas en torno a los datos y la estadística, y el segundo comprende el tratamiento y difusión de datos sobre violencias. La revisión de estos cuerpos de literatura nos permitió identificar que las artes pueden ser medios que, junto a los datos, contribuyen a la creación y transformación de realidades. Así mismo, vimos que los datos plantean retos epistemológicos que nos invitan a buscar formas amigables de compartir con personas no especialistas en estadística para dialogar y reflexionar sobre distintas temáticas. No obstante, es preciso ir más allá de la visualización que propone el arte-dato o de las conversaciones que podemos detonar a partir de los datos. Aún existe un vacío en las formas en las que ambos aspectos conllevan a la construcción de conocimientos, especialmente cuando tratamos con información delicada sobre violencias y vulneraciones a derechos de niñas, niños y adolescentes. Asimismo, las experiencias identificadas y los desafíos mencionados alrededor del análisis, uso y difusión de datos abiertos sobre violencias evidencian importantes aprendizajes que pueden fortalecer el ejercicio de investigación y desarrollo de la estrategia pedagógica del uso reflexivo de datos por parte del OBCE de Bogotá. Específicamente, las recomendaciones alrededor de la calidad de datos, el cuidado con el tratamiento ético, la identificación de los sesgos, y la construcción de protocolos para favorecer el intercambio de información entre entidades. De igual forma, es importante mencionar que, además de la experiencia en identificación de datos en la literatura identificada, no se han encontrado experiencias ilustrativas sobre difusión y apropiación de datos específicamente sobre violencias sobre niños, niñas y jóvenes, aunque sí se identifican ejercicios de análisis y construcción de modelos predictivos al respecto del tema.

¿Qué producto final necesitaría recibir por parte de la Subgerencia de Analítica de Datos para solucionar su reto?	Aplicación que permita navegar en un mapa, panel de control (indicadores).
Amplia tu respuesta anterior. Ejemplos: aplicación que permita navegar en un mapa, panel de control, indicador, información en tiempo real, modelo predictivo, datos con valor agregado entre otros.	Para el OBCE es muy clave el desarrollo de ese producto (mapa) de visualización emergente de la investigación, pues será central para los procesos de artificación del dato y construcción de conocimiento sobre convivencia escolar. Las características esperadas del producto están en definición y construcción.
¿Información adicional que usted quiera mencionar?	El Observatorio de Convivencia Escolar elabora un material denominado Bogodatos para la Convivencia Escolar, en el que se presenta un análisis de los reportes del Sistema de Alertas. La versión semestral de 2024 contine una profundización cualitativa en el módulo de abuso y violencias. Invitamos a su revisión: https://oce.educacionbogota.edu.co/sites/default/files/2025-02/boletin2024primersemestreOBCE_0.pdf Asimismo, invitamos a la exploración de los demás materiales elaborados por el Obce: https://oce.educacionbogota.edu.co/observatorio-de-convivencia-escolar
Autorización tratamiento de datos personales. Autorizo expresamente a la Unidad Administrativa Especial de Catastro Distrital (UAECD) para almacenar los datos personales que yo voluntariamente diligencio en la presente encuesta. Entiendo que estos datos serán custodiados y tratados en los términos de la Ley de Protección de Datos Personales y demás disposiciones concordantes. Comprendo que para revocar o modificar esta autorización, debo dirigirme por escrito a la UAECD. La política para el tratamiento de datos personales se encuentra en el siguiente enlace: http://www.catastrobogota.gov.co/contacto/politicas-de-seguridad-privacidad-condiciones-de-uso-y-proteccion-de-datos-personales	Si autorizo

T3 Tabla 3. Ficha Descriptiva del Proyecto Electrolineras – **Universidad Nacional de Colombia**
Fuente: Elaboración propia.

Ficha Descriptiva – Recepción de retos – Subgerencias de Analítica de Datos			
Información de la Entidad			
Entidad/ Organización	Universidad Nacional de Colombia		
Dependencia/área	Grupo de investigación EM&D		
Enlace	**Correo electrónico**	**Teléfono**	**Cargo**
Luis Felipe Díaz Corzo	ldiazco@unal.edu.co		Asistencial
Información del reto			
Defina cuál es el reto (necesidad, problemática o deseo). Escriba su reto usando la siguiente estructura gramatical: cómo + verbo en infinitivo + complemento. Ejemplo: cómo identificar permanentemente las zonas con mayores casos o situaciones de violencia infantil en Bogotá.	Cómo planificar la ubicación de nuevas estaciones de carga para vehículos eléctricos a partir de GIS en Bogotá		
Descripción detallada del reto. Cuéntenos más sobre éste, el contexto del mismo, indicadores actuales si los hay, cómo lo resuelve actualmente. Si es un deseo explique que le gustaría tener, cómo se imagina la solución ideal, otros.	Dado el aumento de vehículos eléctricos en la ciudad, se quiere utilizar información GIS actual de infraestructura local para crear un modelo de analítica de datos que permita determinar posibles ubicaciones para la instalación de nuevas electrolineras.		
¿Por qué es importante para usted este reto y para qué necesita o quiere resolverlo?	Principalmente, de le grupo de investigación se quería una aplicación de relevancia actual que involucrara el uso de GIS. También, fue seleccionado por interés personal.		
¿Considera que la solución a su reto puede apoyarse con analítica de datos? Responda sí, no, depende o supongo y explique su respuesta.	Si, el uso de GIS multicapa en Python involucra analítica de datos.		
¿Qué tipo de analítica de datos considera se debe desarrollar en la solución de su reto?	Prescriptiva - ¿Qué debería hacer?		

¿La solución, resultado o entregable tendría algún impacto para la ciudad? Justifique su respuesta.	Si, puede conllevar al desarrollo de proyectos de electrolineras nuevas que beneficien a los usuarios de vehículos eléctricos y la población en general.
¿A qué objetivo estratégico, programa o artículo de su entidad u organización y/o del Plan Distrital de Desarrollo está relacionado su reto?	Transición energética.
¿Quién(es) serán los beneficiarios de la solución? Puede haber varios usuarios y beneficiarios. Descríbalos brevemente.	Usuarios de vehículos eléctricos y la población en general gracias a la disminución de emisiones contaminantes en la ciudad.
¿Cuáles son las causas del reto (necesidad, problemáticas o deseos)?	Tendencia de aumento del parque vehicular eléctrico en la ciudad.
¿Cuáles son las consecuencias del reto?	Identificar ubicaciones aptas para el desarrollo de proyectos nuevos de electrolineras.
¿Qué información sugiere o puede compartir para profundizar en el entendimiento del reto?	Documentos, manuales, procedimientos, instructivos institucionales
¿Cuéntenos más sobre los detalles de conceptos, temas, documentos, fuentes de información, instructivos o bibliografía que nos sugiere o puede remitirnos al correo electrónico jmonje@catastrobogota.gov.co, para profundizar en el entendimiento de su reto o expectativas de solución?	Se compartirá al correo de los ingenieros de IDECA que nos han estado colaborando, un documento con lo que se ha hecho hasta el momento.
Capacidades actuales	
¿Cuál es el equipo de trabajo del proyecto que trabaja o trabajará conjuntamente en resolver el reto. Especifique para cada uno: 1. Rol principal, 2. Profesión 3. Si es interno o externo, 4. Si tiene conocimiento o no, en analítica de datos.	Personal de IDECA: Ingenieros encargados de brindar la información GIS disponible y hacer seguimiento al proyecto. Externo. Con conocimientos en analítica de datos. Profesor Javier Rosero: Ingeniero que realiza seguimiento al proyecto desde la parte teórica. Interno. Con conocimientos en analítica de datos. Profesor Oscar Perdomo: Ingeniero que realiza seguimiento al proyecto desde la parte técnica. Interno. Con conocimientos en analítica de datos. Estudiante Luis Felipe Diaz: Ingeniero encargado de

	realizar y documentar el proyecto. Interno. Con conocimientos en analítica de datos.
¿Qué datos o información dispone actualmente en su organización que sean importantes y se puedan usar para dar respuesta al reto identificado para generar la solución que se defina? Descríbalos brevemente: 1. Tipo de dato (estructurado, semiestructurado, no estructurado). 2. Tiene restricción o no para su acceso. 3. El origen y 4. Otros.	Electrolinearas existentes en Bogotá: Estructurado. Sin restricción. Obtenido de electromaps.com. Contiene información de su capacidad de potencia, número de cargadores y velocidad de carga. Uso del suelo en Bogotá: Estructurado. Sin restricción. Obtenido de IDECA. Capa de manzanas de Bogotá: Estructurado. Sin restricción. Obtenido de IDECA. Infraestructura eléctrica de media tensión de Bogotá: Estructurado. Con restricción. Obtenido de ENEL CODENSA. Contiene información de apoyos y líneas de media tensión. Subestaciones de alta tensión en Bogotá: Estructurado. Con restricción. Obtenido del Profesor Javier Rosero. Matrices de origen-destino de movilidad en Bogotá: Estructurado. Con restricción. Obtenido de IDECA. Contiene información georreferenciada de los puntos de origen y destino de movilidad por hora y actividad a realizar por los encuestados.
¿Dispone de recursos tecnológicos (Hardware o Software) para usarse durante el proceso de generación de la solución? Descríbalos brevemente.	Si, computador personal con Python.
¿Tiene disponibilidad de recursos económicos para apoyar las actividades o recursos necesarios para llevar a cabo el desarrollo de la solución esperada?	No puedo dar respuesta en este momento;
Actualmente en qué estado de avance o implementación tiene su idea o solución internamente en su entidad	Tengo un prototipo o algo para probar si funciona de manera muy básica sin muchos detalles.
La solución imaginada	
¿Existe alguna solución o prototipo para resolver este reto que usted conozca o sea un referente?	En la literatura hay muchos proyectos para planificación de este equipamiento.

¿Qué producto final necesitaría recibir por parte de la Subgerencia de Analítica de Datos para solucionar su reto?	Modelos basados en algoritmos y programación, información en tiempo real, modelo predictivo.
Amplia tu respuesta anterior. Ejemplos: aplicación que permita navegar en un mapa, panel de control, indicador, información en tiempo real, modelo predictivo, datos con valor agregado entre otros.	Script de programación que permita identificar los puntos exactos donde es viable colocar electrolineras en Bogotá, a partir de la capas de datos estudiadas.
¿Información adicional que usted quiera mencionar?	Sin información
Autorización tratamiento de datos personales. Autorizo expresamente a la Unidad Administrativa Especial de Catastro Distrital (UAECD) para almacenar los datos personales que yo voluntariamente diligencio en la presente encuesta. Entiendo que estos datos serán custodiados y tratados en los términos de la Ley de Protección de Datos Personales y demás disposiciones concordantes. Comprendo que para revocar o modificar esta autorización, debo dirigirme por escrito a la UAECD. La política para el tratamiento de datos personales se encuentra en el siguiente enlace: http://www.catastrobogota.gov.co/contacto/politicas-de-seguridad-privacidad-condiciones-de-uso-y-proteccion-de-datos-personales	Si autorizo

T4 Tabla 4. Ficha Descriptiva del Proyecto Biomasa – **Universidad Nacional de Colombia**
Fuente: Elaboración propia.

Ficha Descriptiva – Recepción de retos – Subgerencias de Analítica de Datos		
Datos generales del solicitante		
Entidad/ Organización	Universidad Nacional de Colombia	
Dependencia/área	Departamento de Ingeniería Eléctrica y Electrónica	
Nombres y apellidos	**Correo electrónico**	**Nivel ocupacional**
Yessica Vanessa Trujillo Ladino	ytrujillol@unal.edu.co	Profesional
Información del reto		
Defina cuál es el reto (necesidad, problemática o deseo). Escriba su reto usando la siguiente estructura gramatical: cómo + verbo en infinitivo + complemento. Ejemplo: cómo identificar permanentemente las zonas con mayores casos o situaciones de violencia infantil en Bogotá.	Cómo priorizar ubicaciones para microhubs de biogás integrando datos oficiales en SIG para aprovechar residuos de podas, plazas de mercado y colegios distritales en Bogotá, minimizando costos logísticos y cumpliendo restricciones ambientales y de conexión a la red eléctrica.	
Descripción detallada del reto. Cuéntenos más sobre éste, el contexto del mismo, indicadores actuales si los hay, cómo lo resuelve actualmente. Si es un deseo explique que le gustaría tener, cómo se imagina la solución ideal, otros.	El reto consiste en pasar de la simple disposición de residuos orgánicos al aprovechamiento energético local en Bogotá, usando información geoespacial oficial para decidir dónde y cómo instalar soluciones cercanas a la fuente. Hoy la ciudad genera más de 6.300 toneladas diarias de residuos sólidos, de las cuales cerca del 60 % es fracción orgánica, aproximadamente 3.700 t/día con potencial de valorización. La mayor parte termina en el relleno sanitario Doña Juana, donde se pierde como recurso y se generan emisiones de metano no controladas. Al mismo tiempo, equipamientos como los colegios distritales requieren energía confiable para usos cotidianos. Tecnologías probadas como la digestión	

	anaerobia permiten transformar 1 tonelada de residuos orgánicos en 100 a 200 m³ de biogás de manera referencial, pero para materializar esa oportunidad es indispensable saber dónde están las fuentes, qué volúmenes se generan y qué tan accesibles son. Indicadores y estado actual. Disponemos de cifras agregadas sobre generación y del ecosistema de datos de IDECA y entidades distritales para mapear vialidad, restricciones y localización de posibles offtakers. Sin embargo, persisten vacíos operativos: planes y cuadrantes de poda no estandarizados como datos abiertos, información dispersa sobre patios de acopio y ausencia de factores locales de generación en colegios por alumno y por estacionalidad académica. En la práctica, la gestión se resuelve principalmente mediante recolección y disposición; no existe un esquema sistemático de aprovechamiento energético integrado a la planeación espacial de la ciudad. Qué se desea y solución ideal. Contar con una metodología basada en SIG que integre datos oficiales y criterios técnicos, logísticos, socioambientales y normativos para: 1. Identificar y cuantificar las fuentes urbanas de biomasa más relevantes, con foco en podas, plazas de mercado y colegios. 2. Evaluar accesibilidad por red vial, restricciones ambientales y proximidad a la red eléctrica para asegurar viabilidad técnica. 3. Priorizar ubicaciones para microhubs y puntos de acopio cercanos a la fuente, reduciendo transporte y pérdidas. 4. Caracterizar la demanda en colegios como offtakers potenciales e impulsar prediseños tecnológicos con balances de masa y energía. 5. Entregar mapas de potencial e idoneidad, listas cortas de sitios y una guía metodológica reproducible, además de un artículo científico que documente resultados y aprendizajes. En síntesis, el reto busca sustituir la disposición en relleno por soluciones de bioenergía de proximidad, apoyadas en datos geoespaciales confiables, para obtener beneficios ambientales, logísticos y energéticos medibles en la ciudad.
¿Por qué es importante para usted este reto y para qué necesita o quiere resolverlo?	Para mí este reto importa porque quiero transformar un residuo que hoy se desperdicia en energía útil para las comunidades, empezando por los colegios, mientras aprendo

	nuevas estrategias y las aplico a un problema real de la ciudad. Quiero pasar del papel a decisiones territoriales claras con datos confiables: ubicar oportunidades, estimar cuánto aporta la biomasa y definir dónde conviene instalar microhubs cercanos a la red. En concreto, busco reducir emisiones, recortar costos de transporte y dejar una metodología sencilla y reproducible que IDECA y las entidades puedan usar, a la vez que fortalezco mis capacidades técnicas y de impacto público con resultados medibles en el territorio.
¿Considera que la solución a su reto puede apoyarse con analítica de datos? Responda sí, no, depende o supongo y explique su respuesta.	Sí. La analítica de datos es clave para este reto porque necesito examinar, limpiar, transformar e interpretar datos geoespaciales y operativos para pasar del diagnóstico a decisiones territoriales concretas.
¿Qué tipo de analítica de datos considera se debe desarrollar en la solución de su reto?	Combinación de varias
¿La solución, resultado o entregable tendría algún impacto para la ciudad? Justifique su respuesta.	Sí. Al aprovechar la biomasa de los residuos para generar energía cerca de la fuente y reducir la disposición en relleno, el proyecto aporta beneficios ambientales, operativos y sociales, especialmente para colegios y equipamientos públicos.
¿A qué objetivo estratégico, programa o artículo de su entidad u organización y/o del Plan Distrital de Desarrollo está relacionado su reto?	Se alinea con la transición energética y la economía circular del Plan Distrital de Desarrollo: aprovecha la fracción orgánica para generar energía y disminuir envío a Doña Juana. Contribuye a acción climática (metano evitado) y a la gestión basada en datos al usar información geoespacial oficial para decidir dónde intervenir. Sectorialmente, fortalece la misión de IDECA y apoya metas de UAESP, SED, JBB y SDA en valorización de orgánicos y eficiencia logística.
¿Quién(es) serán los beneficiarios de la solución? Puede haber varios usuarios y beneficiarios. Descríbalos brevemente.	Los beneficiarios son, principalmente, los colegios distritales como offtakers de energía para cocina, agua caliente e iluminación.
¿Cuáles son las causas del reto (necesidad, problemáticas o deseos)?	Las causas son: alta generación de fracción orgánica que termina en el relleno; datos dispersos y vacíos operativos;

	débil articulación entre oferta de biomasa y demanda útil con incertidumbre en rendimientos; marcos normativos y de gobernanza fragmentados y pocos incentivos a la valorización; percepciones sociales por olores en entornos sensibles; y necesidad de capacidades para aplicar analítica y SIG que permitan decisiones territoriales reproducibles.
¿Cuáles son las consecuencias del reto?	Las consecuencias son: mayor disposición de orgánicos en el relleno con emisiones de metano y pérdida de valor energético; más kilómetros de recolección y costos logísticos altos; servicios menos confiables en colegios que podrían usar energía local; molestias por olores y camiones en barrios; oportunidad económica y de empleo desaprovechada en microhubs y acopio; decisiones públicas con poca base de datos y menor trazabilidad; y retraso en metas de economía circular, acción climática y transición energética del Distrito.
¿Qué información sugiere o puede compartir para profundizar en el entendimiento del reto?	Recursos digitales, audiovisuales, web
¿Cuéntenos más sobre los detalles de conceptos, temas, documentos, fuentes de información, instructivos o bibliografía que nos sugiere o puede remitirnos al correo electrónico jmonje@catastrobogota.gov.co, para profundizar en el entendimiento de su reto o expectativas de solución?	Sin información
Capacidades actuales	
¿Cuál es el equipo de trabajo del proyecto que trabaja o trabajará conjuntamente en resolver el reto. Especifique para cada uno: 1. Rol principal, 2. Profesión 3. Si es interno o externo, 4. Si tiene conocimiento o no, en analítica de datos.	El equipo está conformado por Javier Rosero García, Ph.D. (director del proyecto), ingeniero electricista e investigador del grupo EM&D de la Universidad Nacional de Colombia, interno y con conocimiento en analítica de datos; Yessica Vanessa Trujillo, matemática, responsable de análisis de datos y modelado geoespacial, interna y con conocimiento en analítica de datos; y Juan Galvis, Ph.D., matemático, colaborador en el modelado matemático y validación de supuestos, interno y con conocimiento en analítica de datos.

¿Qué datos o información dispone actualmente en su organización que sean importantes y se puedan usar para dar respuesta al reto identificado para generar la solución que se defina? Descríbalos brevemente: 1. Tipo de dato (estructurado, semiestructurado, no estructurado). 2. Tiene restricción o no para su acceso. 3. El origen y 4. Otros.	Para poda urbana, contamos con el Inventario de arbolado urbano en formato vectorial estructurado y de acceso abierto, con origen en JBB/Datos Abiertos Bogotá; lo usamos como base para estimar la oferta de poda a partir de atributos dendrométricos y de estado. Complementa el ráster de densidad de arbolado (dato semiestructurado, acceso público vía servicio de mapas, origen SIGAU–JBB), útil para identificar concentraciones territoriales. Sumamos el Sistema distrital de parques (vectorial estructurado, abierto, origen IDECA/Datos Abiertos). Para colegios distritales, disponemos de la localización oficial de sedes educativas (vectorial estructurado, abierto, origen SED/Datos Abiertos), que nos permite mapear puntos de generación; la matrícula por sede (tabular estructurado, abierto, origen SED/Datos Abiertos), que usamos como proxy para estimar residuos alimentarios; y el consumo de energía por sede (tabular estructurado, abierto, origen SED/Datos Abiertos), con el que relacionamos la oferta de biomasa con usos energéticos in situ. También usamos la capa de uso del suelo/área de actividad (vectorial estructurado, abierto, origen IDECA/Datos Abiertos) para verificar compatibilidad urbanística. Para plazas de mercado y centrales de abasto, contamos con la ubicación de plazas distritales (vectorial estructurado, abierto, origen IDECA/Datos Abiertos), insumo primario para identificar fuentes concentradas y trazables. Como capas transversales para los tres casos, usamos la malla vial integral (vectorial/red estructurado, abierto, origen IDECA/Datos Abiertos) para modelar accesibilidad, tiempos de viaje y costos logísticos; las restricciones ambientales como ronda hidráulica, sistema hídrico, amenaza por inundación y amenaza por movimiento en masa (vectorial estructurado, abierto, origen IDECA/IDIGER/SDA) para definir exclusiones y umbrales de sostenibilidad territorial; y la información de vientos/rosas de viento (dato semiestructurado, acceso público de consulta, origen RMCAB) para orientar ubicaciones que minimicen molestias por olores hacia receptores sensibles.
¿Dispone de recursos tecnológicos (Hardware o Software) para usarse	Sí. Contamos con hardware de cómputo portátil y de escritorio para trabajo geoespacial (CPU multinúcleo, 16 GB RAM y GPU

durante el proceso de generación de la solución? Descríbalos brevemente.	básica) y almacenamiento externo para respaldos. En software, usamos QGIS para análisis y cartografía, Python (pandas, geopandas, shapely) para limpieza y modelado.
¿Tiene disponibilidad de recursos económicos para apoyar las actividades o recursos necesarios para llevar a cabo el desarrollo de la solución esperada?	No tengo recursos económicos
Actualmente en qué estado de avance o implementación tiene su idea o solución internamente en su entidad	Tengo un prototipo o algo para probar si funciona de manera muy básica sin muchos detalles.
La solución imaginada	
¿Existe alguna solución o prototipo para resolver este reto que usted conozca o sea un referente?	Sin información
¿Qué producto final necesitaría recibir por parte de la Subgerencia de Analítica de Datos para solucionar su reto?	Datos o análisis de datos para una investigación académica
Amplia tu respuesta anterior. Ejemplos: aplicación que permita navegar en un mapa, panel de control, indicador, información en tiempo real, modelo predictivo, datos con valor agregado entre otros.	Nos hacen falta, principalmente, datos operativos y de detalle: en poda urbana, un dataset descargable y estandarizado de cuadrantes y programación (geometrías, frecuencia y últimas/próximas intervenciones), la ubicación y capacidad de patios de acopio/CTA, la estacionalidad mensual por localidad o cuadrante y la composición de la poda (humedad, C/N, sólidos volátiles); en colegios, el factor de generación de residuos alimentarios (kg/alumno·día) por nivel y jornada; en plazas de mercado, el volumen diario de orgánicos por plaza con variación temporal, la composición biodegradable vs impropios y las ventanas horarias con puntos internos de acopio.
¿Información adicional que usted quiera mencionar?	Sin información
Autorización tratamiento de datos personales.	Si autorizo

Autorizo expresamente a la Unidad Administrativa Especial de Catastro Distrital (UAECD) para almacenar los datos personales que yo voluntariamente diligencio en la presente encuesta. Entiendo que estos datos serán custodiados y tratados en los términos de la Ley de Protección de Datos Personales y demás disposiciones concordantes. Comprendo que para revocar o modificar esta autorización, debo dirigirme por escrito a la UAECD. La política para el tratamiento de datos personales se encuentra en el siguiente enlace: http://www.catastrobogota.gov.co/contacto/politicas-de-seguridad-privacidad-condiciones-de-uso-y-proteccion-de-datos-personales	

T5 Tabla 5. Ficha Descriptiva del Proyecto Paneles Solares – **Universidad Nacional de Colombia**
Fuente: Elaboración propia.

Ficha Descriptiva – Recepción de retos – Subgerencias de Analítica de Datos		
Información de la Entidad		
Entidad/ Organización	UNIVERSIDAD NACIONAL DE COLOMBIA	
Dependencia/área	INGENIERÍA	
Enlace	**Correo electrónico**	**Nivel ocupacional/cargo**
ROLAND STEVEND PORTILLA GARCIA	RPORTILLA@UNAL.EDU.CO	ESTUDIANTE INVESTIGADOR
Información del reto		
Defina cuál es el reto (necesidad, problemática o deseo). Escriba su reto usando la siguiente estructura gramatical: cómo + verbo en infinitivo + complemento. Ejemplo: cómo identificar permanentemente las zonas con mayores casos o situaciones de violencia infantil en Bogotá.	Cómo evaluar la viabilidad de proyectos de generación fotovoltaica en Bogotá a través de una metodología replicable basada en información georreferenciada.	
Descripción detallada del reto. Cuéntenos más sobre éste, el contexto del mismo, indicadores actuales si los hay, cómo lo resuelve actualmente. Si es un deseo explique que le gustaría tener, cómo se imagina la solución ideal, otros.	El reto consiste en evaluar y priorizar, con criterios técnicos reproducibles, el potencial de generación fotovoltaica distribuida en Bogotá (especialmente en cubiertas urbanas), integrando datos georreferenciados (catastro/uso del suelo/morfología), datos meteorológicos y modelos de radiación para soportar decisiones de inversión y despliegue. El contexto está marcado por el aumento sostenido del consumo eléctrico nacional y la necesidad de fortalecer la seguridad energética urbana, junto con los compromisos de mitigación climática. Indicadores actuales :	

	- Crecimiento del consumo eléctrico nacional 2012–2023. - En Bogotá, irradiancia estimada alrededor de 4,0377 kWh/m²·día (competitiva para FV, aunque por debajo del promedio nacional). - Meta distrital de reducción de GEI al 2027. - Dos casos de estudio aplicados en colegios de Bogotá: dimensionamientos del orden de 72,6–95,0 kWp, con resultados financieros robustos (VPN positivo, TIR ~19%, payback ~5 años). Actualmente se prioriza las evaluaciones puntuales, no estandarizadas y con déficit de integración (radiación + meteorología + inventario morfológico de cubiertas) para la resolución del problema, lo que dificulta pasar de "posibilidades técnicas" a una priorización basada en evidencia. Y la solución ideal sería una herramienta/metodología "de punta a punta" que produzca mapas de viabilidad, ranking de predios/sectores, estimaciones de energía, autoconsumo, excedentes, e indicadores económico-ambientales (VAN/TIR/LCOE/CO_2 evitado), con trazabilidad y capacidad de réplica en distintos tipos de edificaciones.
¿Por qué es importante para usted este reto y para qué necesita o quiere resolverlo?	Porque habilita una toma de decisiones pública y técnica para acelerar la transición energética urbana: permite reducir incertidumbre, evitar inversiones mal focalizadas, y alinear proyectos FV con objetivos de seguridad energética y descarbonización del Distrito. Lo que como ingeniero considero que es fundamental en el proceso de transición energética hacia la generación distribuida y descarbonización de la matriz energética del país.

¿Considera que la solución a su reto puede apoyarse con analítica de datos? Responda sí, no, depende o supongo y explique su respuesta.	Sí. El reto, por definición, depende de analítica de datos porque requiere integrar fuentes heterogéneas, transformar datos espaciales en variables de interés y utilidad para el proyecto como pendiente, orientación, sombra y POA, y finalmente generar indicadores accionables para priorización y decisión de inversión, siempre fundamentado en el contexto del proyecto que por depender de las características morfológicas del entorno esta estrictamente realcionado a la infromacion georreferenciada.
¿Qué tipo de analítica de datos considera se debe desarrollar en la solución de su reto?	Prescriptiva - ¿Qué debería hacer?
¿La solución, resultado o entregable tendría algún impacto para la ciudad? Justifique su respuesta.	Sí. La solución aporta impacto directo en: - Planeación energética urbana (priorizar dónde conviene instalar proyectos de generación fotovoltaica). - Mitigación climática al cuantificar contribución a reducción de emisiones. - Equidad energética, al facilitar modelos replicables para infraestructura pública y/o esquemas comunitarios. Por ejemplo, la ciudad ya impulsa programas de generación solar comunitaria con beneficios en factura para hogares vulnerables, lo que muestra pertinencia y escalabilidad del reto.
¿A qué objetivo estratégico, programa o artículo de su entidad u organización y/o del Plan Distrital de Desarrollo está relacionado su reto?	El proyecto se conecta con el eje de acción climática y metas de reducción de GEI para 2027, donde la FV urbana actúa como habilitador técnico para mitigación y resiliencia.
¿Quién(es) serán los beneficiarios de la solución? Puede haber varios usuarios y beneficiarios. Descríbalos brevemente.	Entidades distritales de planeación, medio ambiente, hábitat, educación e infraestructura, entro otros que requieren priorización técnica y soporte de inversión. Instituciones públicas como colegios al poder ver la posible reducción de costos energéticos y mejora de sostenibilidad al evaluar su potencial de generación y contar con un criterio técnico en el desarrollo de posibles proyectos de generación.

	Ciudadanía: beneficios indirectos por mitigación en la concentración de GEI, tecnificación y respaldo de los sistemas de generación y potenciales esquemas comunitarios de generación.
¿Cuáles son las causas del reto (necesidad, problemáticas o deseos)?	Brecha de información integrada en Bogotá: ausencia de integración local de radiación, meteorología de alta resolución e inventario morfológico de cubiertas (pendiente/orientación/sombra). Alta variabilidad atmosférica local (nubosidad) y complejidad urbana, que exige modelos y control de calidad de datos para no sesgar resultados. Necesidad de trazabilidad: sin una metodología replicable, los estudios quedan como “casos aislados” difíciles de escalar.
¿Cuáles son las consecuencias del reto?	Priorización subóptima de inversiones FV (se instalan proyectos donde no maximizan beneficios). Retrasos en metas climáticas/energéticas por falta de evidencia cuantitativa local. Oportunidades perdidas de ahorro y eficiencia en infraestructura pública.
¿Qué información sugiere o puede compartir para profundizar en el entendimiento del reto?	Documentos, manuales, procedimientos, instructivos institucionales
¿Cuéntenos más sobre los detalles de conceptos, temas, documentos, fuentes de información, instructivos o bibliografía que nos sugiere o puede remitirnos al correo electrónico jmonje@catastrobogota.gov.co, para profundizar en el entendimiento de su reto o expectativas de solución?	-Como parte del desarrollo del proyecto se podría compartir un documento tipo articulo con el planteamiento metodológico completo y su anexo de implementación en dos colegios de la ciudad de Bogotá. - Como insumo del articulo también se cuenta con las series meteorológicas de irradiancia utilizadas desde la base de datos de la estación meteorológica CATAM u otras con diccionario de datos y periodo.

Capacidades actuales	
¿Cuál es el equipo de trabajo del proyecto que trabaja o trabajará conjuntamente en resolver el reto. Especifique para cada uno: 1. Rol principal, 2. Profesión 3. Si es interno o externo, 4. Si tiene conocimiento o no, en analítica de datos.	Roland Portilla Garcia: estudiante investigador de la Universidad Nacional de Colombia con conocimineto en el manejo de bases de datos y programacion. Javiel Albeiro Rosero: Profesor asesor de la Universidad Nacional de Colombia y lider del proyecto. Equipo de IDECA: equipo de apoyo en el desarroyo del proyecto
¿Qué datos o información dispone actualmente en su organización que sean importantes y se puedan usar para dar respuesta al reto identificado para generar la solución que se defina? Descríbalos brevemente: 1. Tipo de dato (estructurado, semiestructurado, no estructurado). 2. Tiene restricción o no para su acceso. 3. El origen y 4. Otros.	1. Información catastral georreferenciada (predios/edificaciones/cubiertas) - Tipo: estructurado (atributos), semiestructurado (servicios web), geoespacial vectorial. - Restricción de acceso: parcial: el documento usa servicios abiertos (ArcGIS de Catastro Bogotá) para coordenadas, geometrías y atributos básicos. - Origen: Catastro Bogotá (servidor ArcGIS). - Otros: base espacial para validar emplazamiento, extensión de cubiertas y restricciones físicas. 2. Series temporales de irradiancia / meteorología (CATAM – El Dorado) - Tipo: estructurado (series horarias/diarias). - Restricción: usualmente baja-media (según convenio/fuente), en el documento se “descarga y procesa” para el análisis. - Origen: base de datos estación meteorológica CATAM. - Otros: requiere depuración y control de calidad (QA/QC) de atípicos. 3. Topografía urbana / DSM-DTM / LiDAR (si existe en la entidad) - Tipo: no estructurado/semiestructurado (nubes de puntos), estructurado (raster). - Restricción: media-alta (por licencia/cobertura/actualización). - Origen: levantamientos/insumos urbanos. - Otros: útil para sombras, pendientes y orientación; el

	documento reconoce retos de cobertura y costos en enfoques con LiDAR. 4. Perfiles de carga/consumo (por tipología o por sede) - Tipo: estructurado. - Restricción: media-alta (datos sensibles/contratos). - Origen: facturación/gestión energética institucional. - Otros: el documento los contempla como insumo de metodologías híbridas energético-económicas. 5. Datos técnicos y financieros del sistema FV (CAPEX/OPEX, degradación, parámetros de módulo/inversor) - Tipo: estructurado. - Restricción: baja-media (según si son internos o de mercado). - Origen: cotizaciones, históricos, catálogos, bases de costos. - Otros: necesarios para escenarios, sensibilidad e indicadores (VAN/TIR/LCOE).
¿Dispone de recursos tecnológicos (Hardware o Software) para usarse durante el proceso de generación de la solución? Descríbalos brevemente.	Software: el documento implementa la metodología mediante un programa en Python, integrando consulta catastral, procesamiento de irradiancia, modelación y dimensionamiento. Herramientas GIS/modelos: se menciona un enfoque inspirado en ArcGIS Solar Analyst y, como referencia metodológica, herramientas como r.sun / PVGIS / PVSyst. Hardware: para escala ciudad, el documento advierte que ciertos enfoques requieren grandes volúmenes de datos y recursos computacionales avanzados.
¿Tiene disponibilidad de recursos económicos para apoyar las actividades necesarias para llevar a cabo el desarrollo de la solución esperada?	No puedo dar respuesta en este momento
Actualmente en qué estado de avance o implementación tiene su idea o solución internamente en su entidad	Tengo un prototipo o una versión de la solución pero aún no está completamente optimizado o listo para producción.

La solución imaginada	
¿Existe alguna solución o prototipo para resolver este reto que usted conozca o sea un referente?	Sí. actualmente existen diferentes herramientas de modelamiento de sistemas fotovoltaicos: - ArcGIS Solar Analyst / Solar Radiation tools: la metodología implementada se inspira en este enfoque para descomponer radiación directa/difusa e integrar la bóveda celeste. - PVGIS (JRC – Comisión Europea): plataforma para estimar radiación y desempeño FV a escala geográfica. - r.sun (GRASS GIS): modelo para calcular irradiancia/irradiación considerando topografía/sombras, usado como referente metodológico. - PVsyst: software de simulación FV de mayor detalle (pérdidas térmicas, configuración eléctrica, etc.). Entre otros
¿Qué producto final necesitaría recibir por parte de la Subgerencia de Analítica de Datos para solucionar su reto?	Modelos basados en algoritmos y programación, información en tiempo real, modelo predictivo.
Amplia tu respuesta anterior. Ejemplos: aplicación que permita navegar en un mapa, panel de control, indicador, información en tiempo real, modelo predictivo, datos con valor agregado entre otros.	Para que la solución sea verdaderamente útil y escalable en el contexto distrital, el producto final ideal debería integrar una herramienta que permita navegar la información en un mapa y, al mismo tiempo, visualizar indicadores clave en un panel de control, de modo que la priorización de cubiertas y zonas con mayor potencial fotovoltaico se base en evidencia verificable y comparable entre territorios. No obstante, para garantizar que estos resultados no dependan de análisis puntuales o difíciles de replicar, se requiere que la Subgerencia de Analítica de Datos entregue un modelo basado en algoritmos y programación que automatice la integración, depuración, transformación y análisis de datos georreferenciados, meteorológicos y energéticos, asegurando trazabilidad metodológica, control de calidad y capacidad de actualización. Adicionalmente, contar con información en tiempo real (o al menos con actualizaciones periódicas automatizadas) permitiría incorporar variaciones operativas relevantes, como cambios en condiciones atmosféricas, actualizaciones de bases geográficas o variaciones en consumo cuando aplique, evitando que

	los resultados queden rápidamente desactualizados. Finalmente, un modelo predictivo sería necesario para anticipar el desempeño esperado de la generación fotovoltaica bajo escenarios futuros (variabilidad climática, degradación de módulos, incertidumbre tarifaria y demanda), lo cual fortalece la toma de decisiones porque permite evaluar riesgos, comparar alternativas y definir acciones de intervención con mayor certeza técnica y económica.
¿Información adicional que usted quiera mencionar?	1. La calidad y resolución de los insumos condiciona la precisión de la propuesta, enfoques 3D con LiDAR aumentan exactitud para sombras/orientación, pero pueden tener cobertura incompleta/no actualizada y costos altos dependiendo de las herramientas utilizadas para su aprovechamiento, por ejemplo acceso a software de pago como ArcGIS pro. 2. Para una adopción escalable, conviene un enfoque por fases: (i) preselección con capas SIG + climatología; (ii) refinamiento en zonas prioritarias con insumos 3D donde existan. 3. El prototipo descrito ya integra consulta a Catastro Bogotá (ArcGIS) y series de irradiancia CATAM, lo que da una base realista para pasar a un producto institucional.
Autorización tratamiento de datos personales. Autorizo expresamente a la Unidad Administrativa Especial de Catastro Distrital (UAECD) para almacenar los datos personales que yo voluntariamente diligencio en la presente encuesta. Entiendo que estos datos serán custodiados y tratados en los términos de la Ley de Protección de Datos Personales y demás disposiciones concordantes. Comprendo que para revocar o modificar esta autorización, debo dirigirme por escrito a la UAECD. La política para el tratamiento de datos personales se encuentra en el siguiente enlace: http://www.catastrobogota.gov.co/contacto/politicas-de-seguridad-privacidad-condiciones-de-uso-y-proteccion-de-datos-personales	Si autorizo

Nota aclaratoria: Todo lo relacionado con el Proyecto de Incubación del IDRD se encuentra en una memoria técnica a parte.

1.3. Análisis de alternativas de solución de la iniciativa

El análisis de alternativas de solución se realiza mediante la aplicación de la Matriz de Priorización, donde se analizaron cada una de las dimensiones; luego, se realizaron mesas de trabajo con las entidades priorizadas "SED e IDRD"; y este informe corresponde al proyecto de incubación titulado **"Estrategia pedagógica de artificación en el uso reflexivo de los datos"** realizado con el Observatorio de Convivencia Escolar (Obce) adscrito a la Secretaría de Educación (SED).

1.4. Concretar la solución

De las mesas de trabajo realizadas con el Equipo del Observatorio de Convivencia Escolar (Obce), adscrito a la Secretaría de Educación (SED) se concretó como solución a la necesidad **"Desarrollar un modelo piloto predictivo multietiqueta"** que, a partir de variables sociodemográficas (como edad, género, localidad y UPZ), permita predecir las características del posible suceso de abuso y violencia que podría experimentar un estudiante, para lo cual se proponen las siguientes etapas dentro de la propuesta de solución

i) Realizar análisis descriptivos a los datos suministrado;

ii) Procesamiento de datos;

iii) Implementación de la metodología.

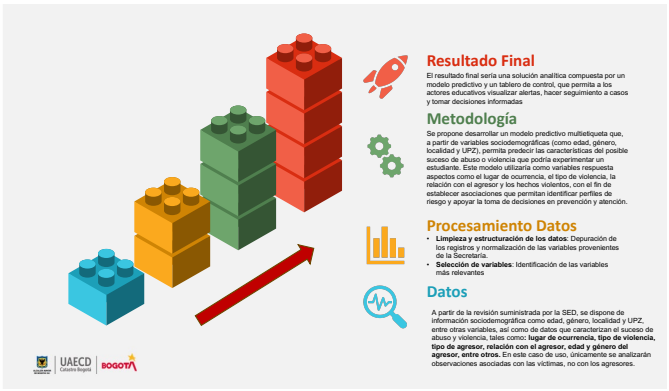
iv) Resultado final

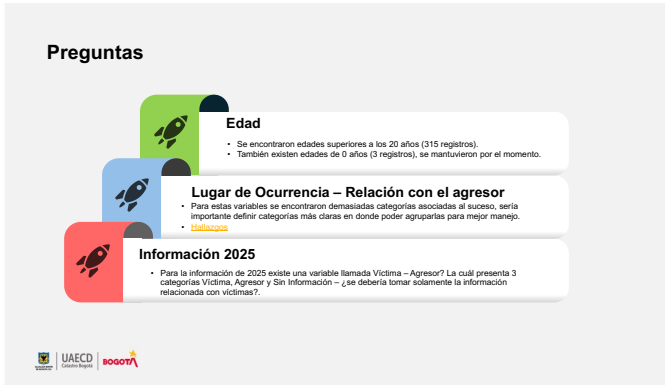
Estas etapas serán desarrolladas en el numeral **1.7: Solucionar la iniciativa.**

1.5. Avalar la idea

En reunión virtual, la propuesta de solución fue presentada al equipo del Obce, así:







1.6. Identificación de Equipos para desarrollar la iniciativa

La Estrategia pedagógica de artificación en el uso reflexivo de los datos como iniciativa de proyecto de incubación se desarrolló con los equipos de la Secretaría de Educación por medio del Observatorio de Convivencia Escolar, la Universidad la Gran Colombia, y la Subgerencia de Analítica de Datos quienes conformarán el Equipo de Trabajo así:

T6 Tabla 6. Equipo de Trabajo para desarrollar el proyecto de incubación
Fuente: Elaboración propia.

Equipo de Trabajo para desarrollar la Iniciativa		
Observatorio de Convivencia Escolar	**Universidad La Gran Colombia**	**Subgerencia de Analítica de Datos**
Jefe Obce Edwin Alberto Ussa Cristiano	**Coordinadora Docente** Nathalie Grajales Olarte.	**Subgerente:** Diego Ricardo Ibarra Rodríguez.
Líder Equipo Contratista Ángela Viviana Roa Ruiz	**Docente** Luz Elena Tinoco Robledo	**Profesional Especializado:** Martha Nubia Rojas Guerrero.
Profesional Especializado Constratista: Dina Luz Riaño Cárdenas		**Profesional Especializado Contratista:** Lilian Daniela Beltrán B.

Adicionalmente, se identificaron otros equipos que puedan estar interesados en desarrollar otros componentes que sirvan para madurar o acelerar esta prueba inicial durante las fases futuras de esta iniciativa u otras de las identificadas en la sección 1.1. de este documento.

1.7. Solucionar la Iniciativa

La solución de la iniciativa comienza con el análisis de los datos suministrados por el Obce a través del enlace público: https://oape.educacionbogota.edu.co/ con lo cual se realizó lo siguiente:

Selección de la Data:

- La base de datos empleada contiene información correspondiente a los años 2014 hasta 2025.
- La información de 2025 cuenta con una variable llamada Víctima – Agresor, la cual presenta tres dominios: -Víctima -Agresor -Sin información; por lo tanto, se tomaron solamente los registros asociados a la categoría de "víctimas".
- Sin embargo, para los datos anteriores a 2025 no se cuenta con este atributo; por lo tanto, se tomó una ventana de datos desde el 2019 hasta el 2025.

Limpieza

- Se tomaron edades entre los 3 y 20 años, omitiendo registros con edades atípicas.
- Se tomaron registros a partir del año 2019 en Adelante.

Datos finales a utilizar

- Una vez realizados los ajustes mencionados anteriormente se obtiene una base preliminar con un total de 113.375 registros.
- Se dispone de información sociodemográfica como edad, género, localidad, entre otras variables, así como de datos que caracterizan el suceso de abuso y violencia, tales como**: tipo de violencia, tipo de agresor, edad y género del agresor.**

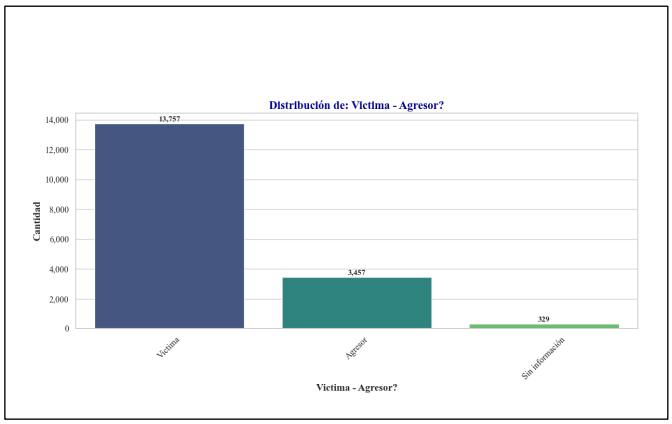
T7 Tabla 7. Variables Explicativas y Variables Respuesta
Fuente: Elaboración propia.

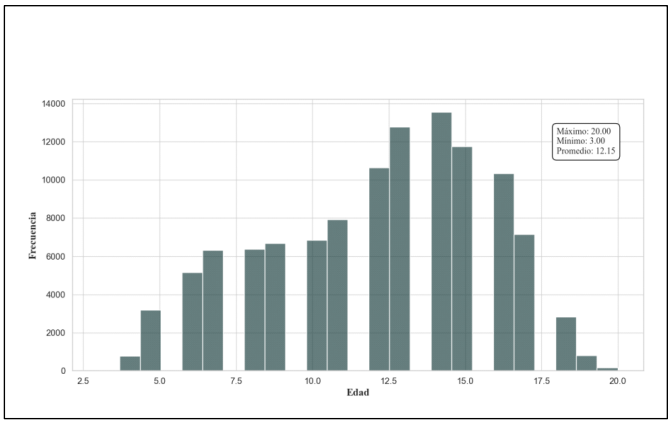
Variables Explicativas	Variables Respuesta
Localidad	Tipo de violencia
Género	Tipo de agresor
Nivel Académico	
Clasificación Edad	
Jornada	

i) Análisis descriptivos de variables – Univariado -

La propuesta presentada se sustentó con base en los análisis descriptivos realizados a la data establecida para la ventana de tiempo 2019 hasta 2025.

G1 *Gráfica 1. Análisis inicial de las variables*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton



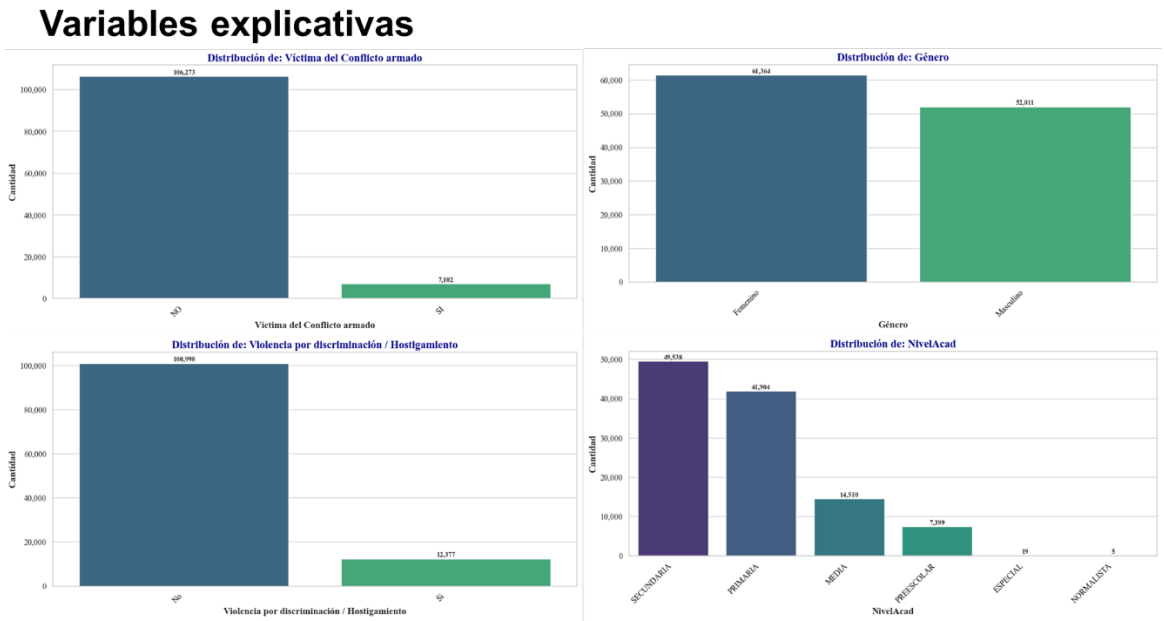


Adicionalmente, se tomaron edades entre los 3 y 20 años, omitiendo registros con edades atípicas. Una vez realizados estos ajustes se obtiene una base preliminar con un total de 113.375 registros.

Finalmente, se dispone de información sociodemográfica como edad, género, localidad, entre otras variables, así como de datos que caracterizan el suceso de abuso y violencia, tales como: tipo de violencia, tipo de agresor, edad y género del agresor.

G2 *Gráfica 2. Variables explicativas*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton

De acuerdo con las variables explicativas consideradas para el modelo, se realizaron los siguientes análisis:

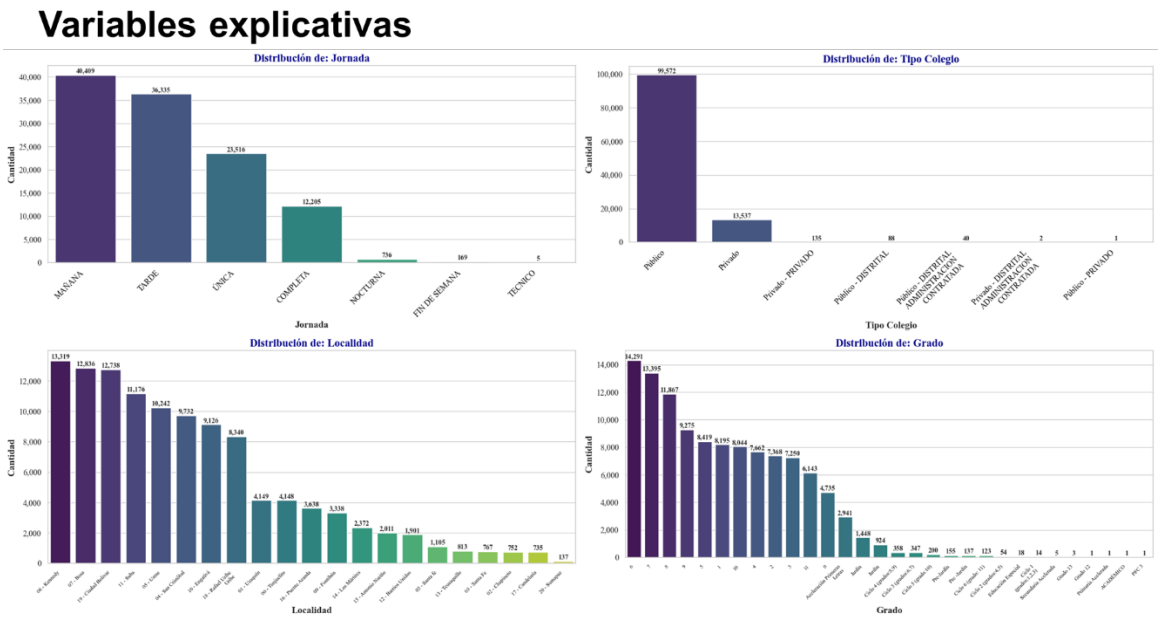


En cuanto a la distribución de las Víctimas del Conflicto armado, se observa **(Gráfica 2)** que la mayoría no fueron víctimas del conflicto.

Respecto al tipo de violencia por discriminación/hostigamiento, se observa que la mayoría de los estudiantes no han sufrido este tipo de violencia.

Finalmente, en cuanto a la distribución por nivel académico, se observa que la mayor cantidad de eventos se presentan en Secundaria, seguido de Primaria.

G3 *Gráfica 3. Análisis inicial de las variables (por jornada, localidad, tipo colegio, grado)*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton



En cuanto a la variable explicativa "Jornada," la mayor cantidad de eventos se presentan en la jornada de la mañana, seguido de la tarde.

Para la variable "Tipo de colegio" se observa una diferencia considerable entre el Colegio Público y el Privado.

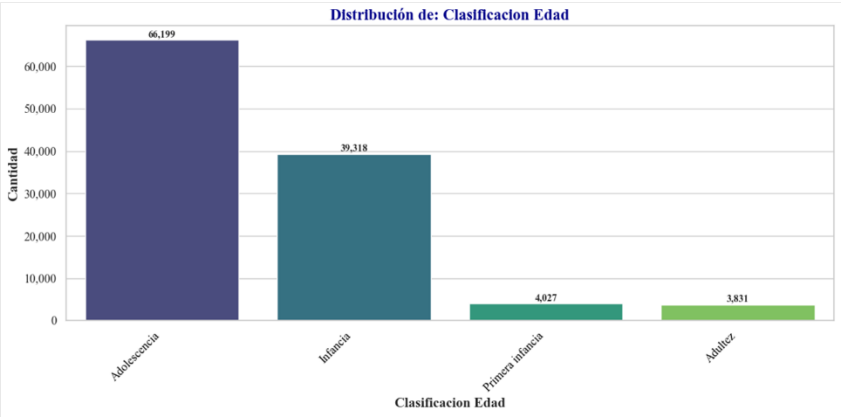
Respecto a la variable "Localidad", las localidades de Kennedy, Bosa, Ciudad Bolívar y Suba, retienen la mayor cantidad de casos.

Esta información de grado se concilia con la distribución "Nivel académico" **(Gráfica 3)**, porque hay correspondencia entre ambas.

G4 *Gráfica 4. Análisis inicial – Variable explicativa Edad*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton

Variables explicativas

- **Primera infancia** 0 a 5 años
- **Infancia** 6 a 11 años
- **Adolescencia** 12 a 17 años
- **Adultez** + 18 años



La variable edad se categorizó de acuerdo con las tipologías enviadas por la SED **(Anexo 1)**, así:

-Primera infancia: 0 -5 años

-Infancia: 6 a 11 años

-Adolescencia: 12 a 17 años

-Adultez: Mayores a 18 años

Por lo anterior, se observa que, en las etapas de la vida, la mayor cantidad de eventos se presentan en la Adolescencia, seguido por la Infancia.

G5 *Gráfica 5. Análisis Variables Respuesta*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton



Existe la posibilidad de incluir variables como Lugar de ocurrencia y Relación con el agresor.

Recordemos que las variables que se determinaron como "Variables Respuesta" son:
-Tipo de violencia
-Tipo de agresor
-Relación con el agresor
-Lugar de ocurrencia

Las anteriores variables permitieron analizar temas como como el género del agresor y el tipo de agresor; sin embargo, se considera de mayor relevancia enfocarse en Tipo de violencia y en Tipo de agresor.

En cuanto a Tipo de violencia, se observa que la violencia sexual y la física son las que más se presentan, seguida por la "Violencia por Negligencia". Y en tipo de agresor, la mayoría de los casos se genera por un "Familiar".
Existen dos variables más que son el lugar de ocurrencia del evento y la relación con el agresor, estas variables tienen múltiples categorías y eso se comunicó en

reunión pasada a la Secretaría (Obce), y se llegó al acuerdo de colaborar en reducir las categorías para poderlas incluir dentro del análisis condensando más las categorías.

ii) Procesamiento de datos;

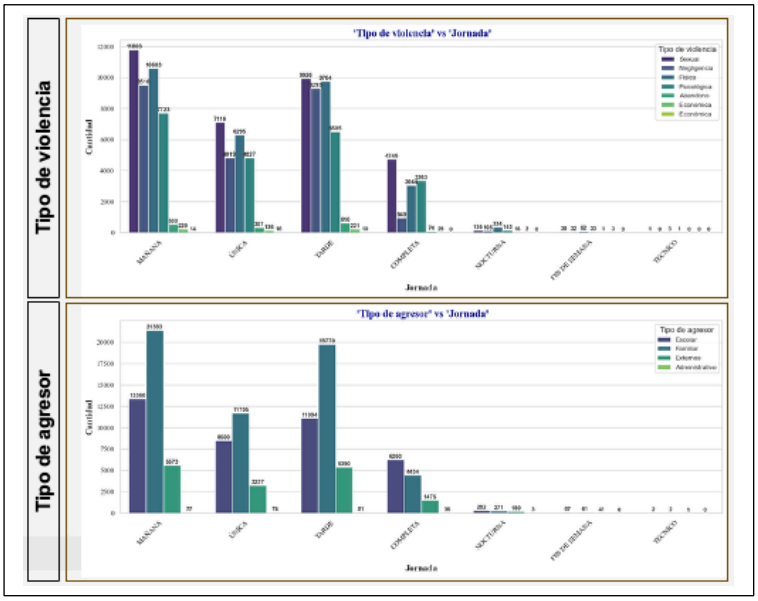
Se realizó un análisis descriptivo Bivariado a la información de acuerdo con las variables seleccionadas como respuesta y las sugeridas como explicativas (Tabla 8).

T8 Tabla 8. Variables Explicativas y Variables Respuesta
Fuente: Elaboración propia.

Variables Explicativas	Variables Respuesta
Localidad Género Nivel Académico Clasificación Edad Jornada	Tipo de violencia Tipo de agresor

Obteniendo los siguientes resultados de análisis:

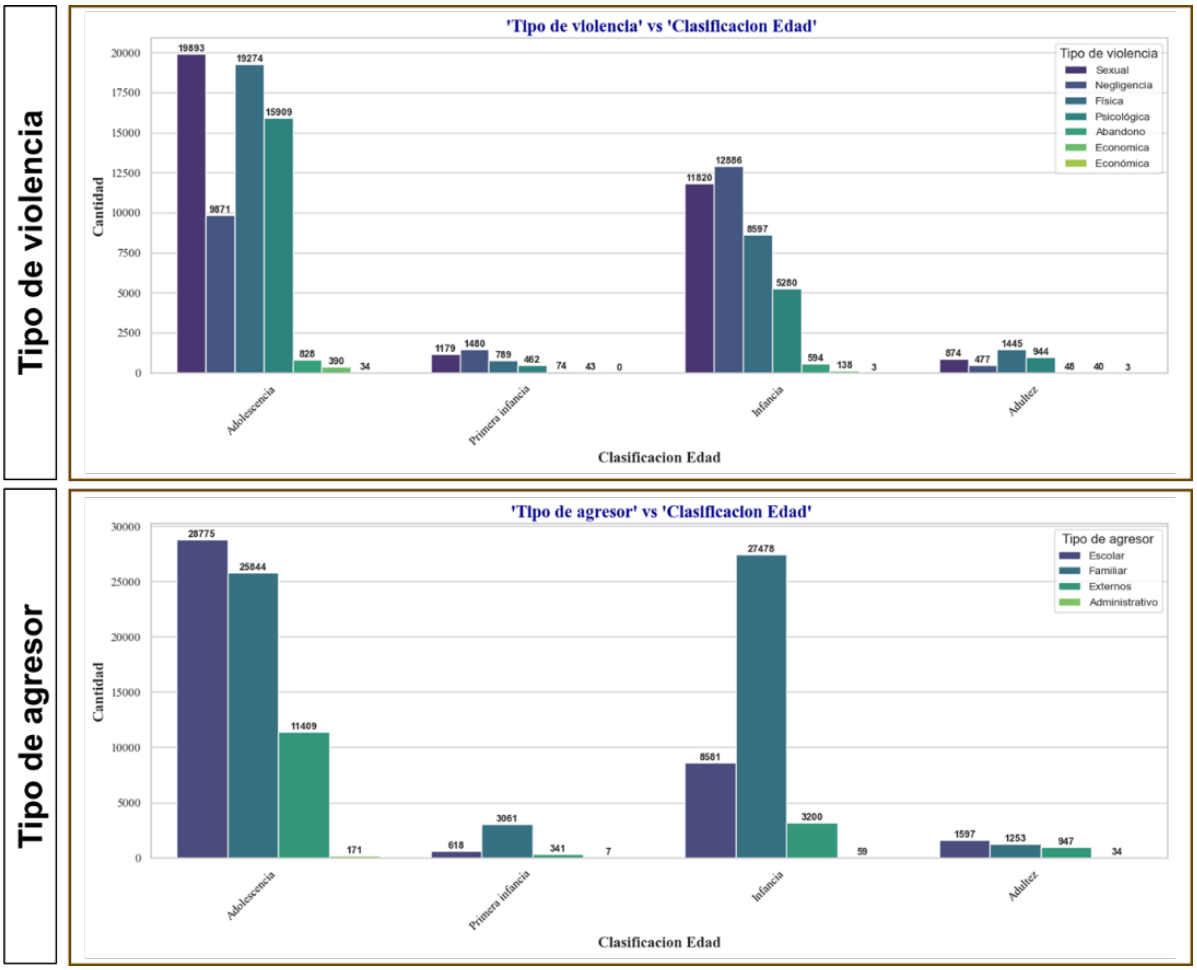
G6 *Gráfica 6. Análisis del tipo de violencia respecto a la jornada*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton



Al analizar el tipo de violencia respecto de la jornada se observa que en todas las jornadas predomina el tipo de violencia sexual, seguida por la violencia física.

Y comparando el tipo de agresor con la jornada se observa que en todas las jornadas predomina el tipo de agresor "Familiar", a excepción de la jornada completa que el tipo de agresor es "Escolar".

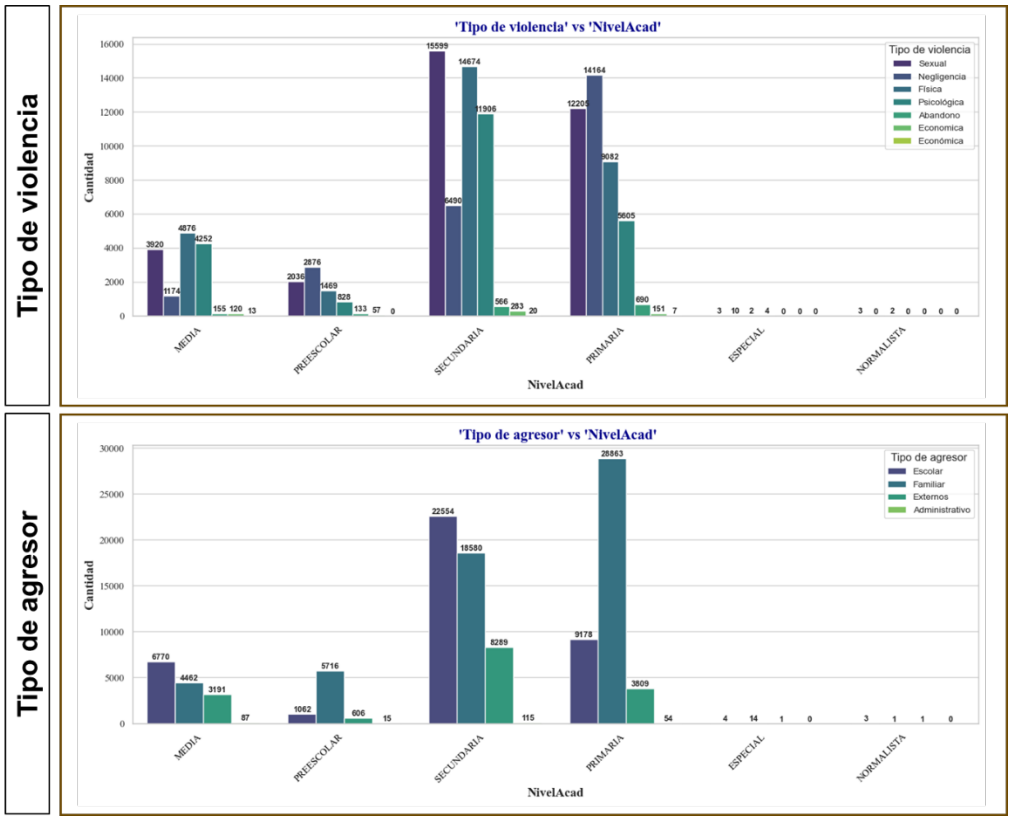
G7 *Gráfica 7. Análisis Tipo de Violencia respecto a los Rangos de Edad*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton



En la **Gráfica 7**, se analiza el tipo de violencia respecto a la clasificación de la edad de los estudiantes, y se observa que en la adolescencia predomina la violencia sexual; en la infancia y primera infancia predomina la violencia por **"negligencia"**; y en la adultez predomina la violencia **"física"**.

Si se revisa por el tipo de agresor, esto es un poco correspondiente, es decir, en la adolescencia el tipo de agresor es mayoritariamente **"escolar"**; mientras que en primera infancia e infancia el agresor es **"familiar"**.

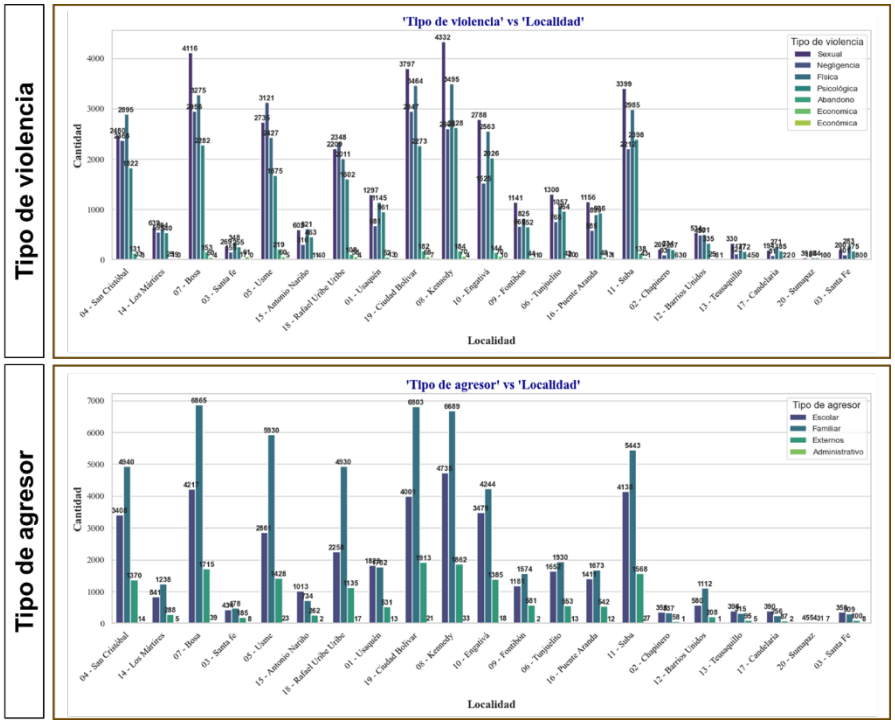
G8 *Gráfica 8. Análisis Tipo de Violencia respecto al Nivel Académico*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton



En el análisis del **Tipo de violencia respecto al Nivel Académico**, se observa que en secundaria es donde se da una mayor cantidad de violencia "sexual", seguido por la violencia física; en cambio, en Primaria la violencia predominante es la "negligencia".

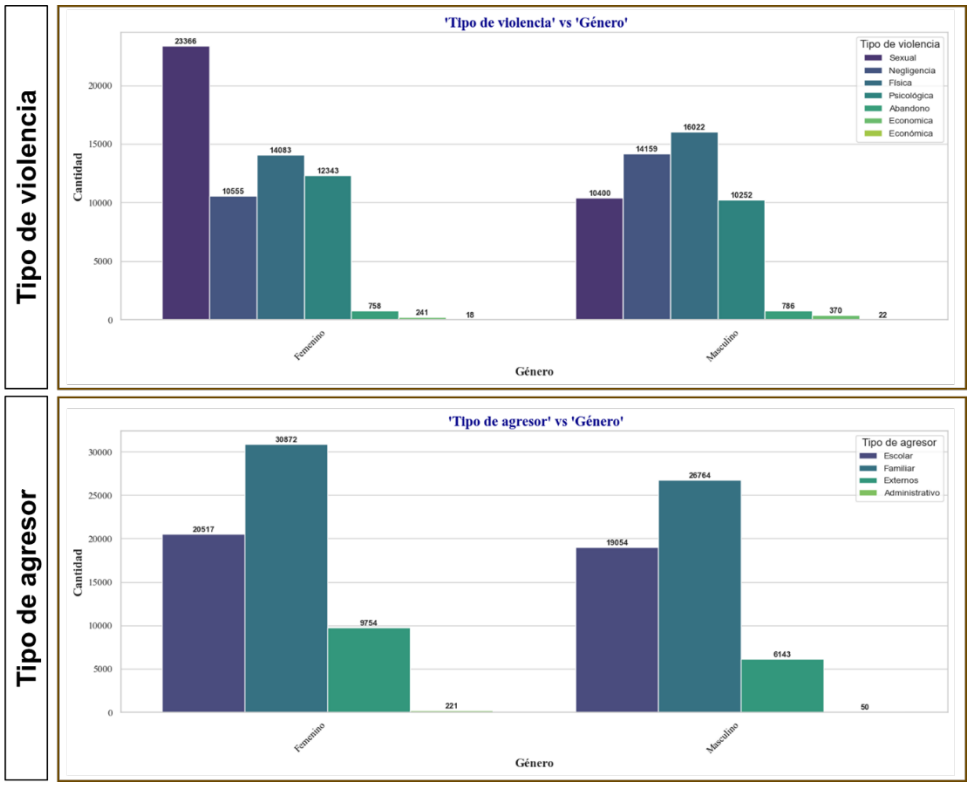
En el análisis de **Tipo de agresor Vs Nivel Académico**, se observa que en primaria el tipo de agresor tiende a ser un familiar; mientras que en secundaria el tipo de agresor es escolar.

G9 *Gráfica 9. Análisis Tipo de Violencia respecto a la Localidad*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton



En el análisis del Tipo de Violencia respecto a la Localidad, se observa que las localidades que agrupan la mayor cantidad de casos son: Bosa, Ciudad Bolívar, Kennedy y Suba; y el tipo de violencia predominante es "sexual", seguida de la violencia "física"; sin embargo, en San Cristóbal la violencia física es la predominante.
Al revisar por Tipo de Agresor, se observa que mayoritariamente corresponde a un "Familiar" en la mayor parte de las localidades.

G10 *Gráfica 10. Análisis Tipo de Violencia respecto al Género*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton



Si se discrimina por género se observa que respecto al **tema femenino (las mujeres)** la "violencia sexual" es la predominante; mientras en el **género masculino** la "violencia física" es mayoritaria.
Y en cuanto al Tipo de Agresor respecto al género, se observa que para los dos (femenino y masculino) el tipo de agresor predominante es el "familiar".

iii) Implementación de la metodología.

Se presenta un Mockup como solución al Modelo Predictivo Propuesto ….

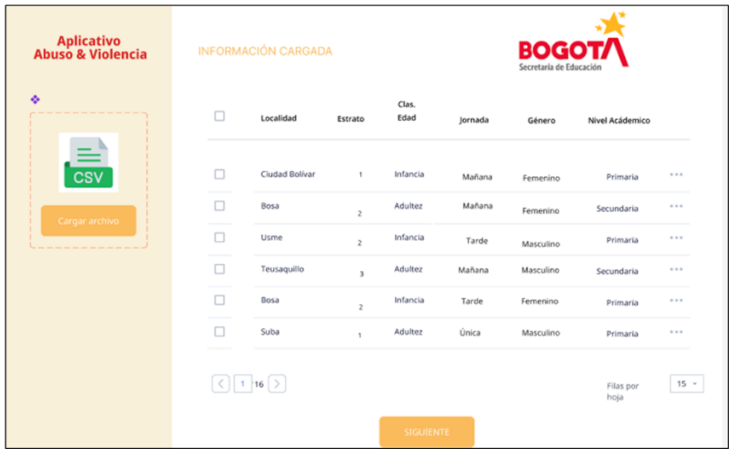
G11 *Gráfica 11. Variables del Modelo (Explicativas y Respuesta)*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton



La variable Estrato no se encuentra incluida en los datos suministrados; asimismo, las variables Relación con el agresor y Lugar de ocurrencia están sujetas a revisión, debido a la gran cantidad de categorías que presentan.

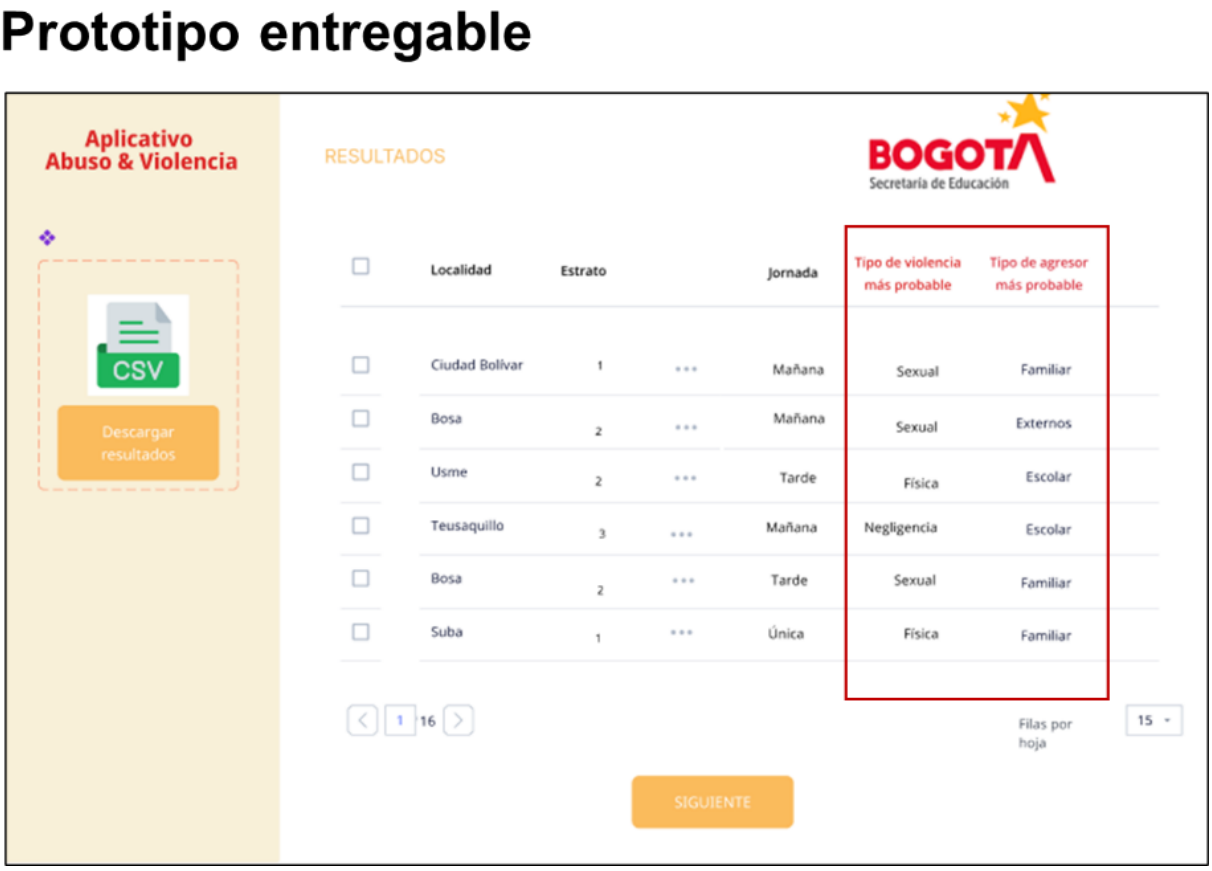
G12 *Gráfica 12. Diseño de Prototipo Entregable*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton

Prototipo entregable



El prototipo (MockUp) representa el aplicativo de lo que se entregaría, y corresponde a una interfaz donde se pueda cargar un archivo con el listado de los estudiantes a los que se desee realizarles el estudio sobre la posibilidad de tener algún tipo de violencia con las variables explicativas que son las sociodemográficas – Gráfica 12.

G13 *Gráfica 13. Diseño de Prototipo Entregable*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton



Luego de cargado el archivo con los datos de los estudiantes, el aplicativo arrojaría para cada estudiante el tipo de violencia y el tipo de agresor más probable (Gráfica 13).

El modelo asignaría una categoría de violencia y agresor a cada estudiante.

NOTA: Si se obtuviera información de estudiantes sin eventos se podrían determinar los que no presenten ningún riesgo.

iv) Resultado final

El procesamiento de la información y diseño del modelo se realizó a través del planteamiento de tres escenarios por medio de los cuales se probaron 9 experimentos aplicando cuatro modelos de Machine Learning (ML), como se describe a continuación:

Escenario Base: mantiene los datos como vienen. En este escenario se hicieron 4 experimentos, aplicando los siguientes 4 modelos (un experimento por modelo):

-Modelo Logístico

-Modelo Random Forest

-Modelo XGBOOST

-Modelo MLP

Escenario Hiperparámetros: Se realizó un experimento con hiperparámetros.

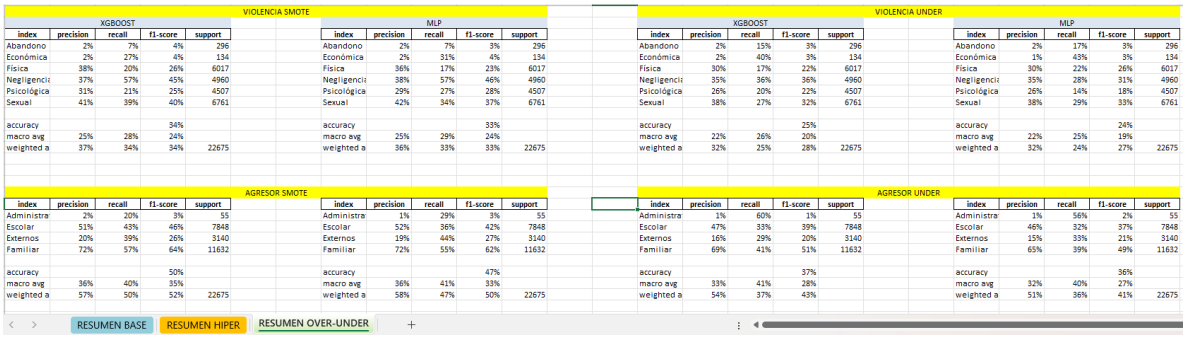
Escenario Over-Under: Se tomaron los modelos XGBOOST Base y MLP Hiperparámetros (por ser los que mejores métricas dieron); y a estos modelos seleccionados se les aplicó "Sobre Muestreo (Smote)" para un equivalente a dos experimentos más; y también, se les aplicó Submuestreo, con lo cual se obtuvieron dos experimentos adicionales.

En total se desarrollaron 9 experimentos aplicando cuatro modelos de Machine Learning (ML).

Detalle de los experimentos realizados

Los 9 experimentos desarrollados se materializan en este archivo de Excel **(Gráfica 14)** soportada con las siguientes tres hojas de cálculo:

G14 *Gráfica 14. Experimentos realizados*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton



Hoja resumen base: se tienen 6 clases de violencias las cuales están desbalanceadas **(Gráfica 15)**, entonces se probaron con los modelos logístico, un Random Forest, un XGBoost, y un Percepto multicapas, y así se obtuvieron las métricas de violencia y agresor.

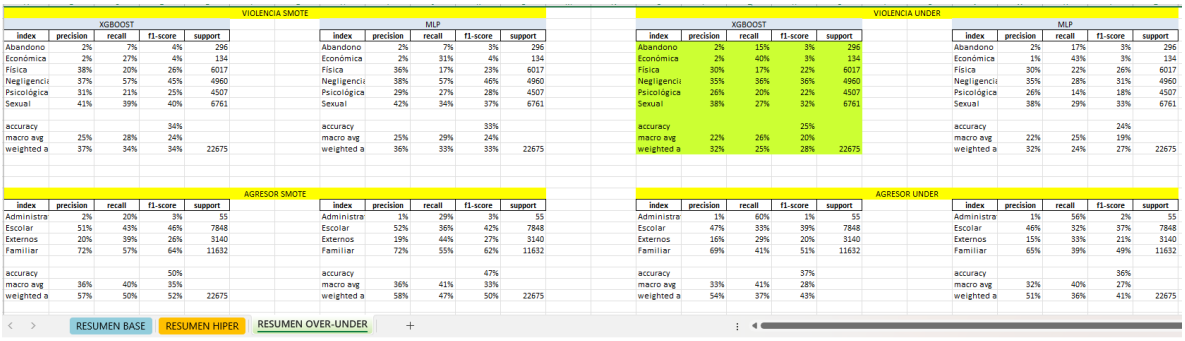
G15 *Gráfica 15. Datos de las violencias desbalanceados*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton

LOGISTIC				
index	**precision**	**recall**	**f1-score**	**support**
Abandono	0%	0%	0%	296
Económica	0%	0%	0%	134
Física	37%	33%	35%	6017
Negligencia	40%	47%	43%	4960
Psicológica	34%	5%	9%	4507
Sexual	39%	63%	48%	6761
accuracy			39%	
macro avg	25%	25%	22%	
weighted avg	37%	39%	35%	22675

Según la **Gráfica 15**, en la columna **support** se almacena la cantidad de individuos que hay en cada clase, las violencias de **Abandono** y **Económica** tiene menor

cantidad de individuos a comparación de las violencias de negligencia, psicológica y sexual, por lo tanto, al entrenar un modelo de Machine Learning con este desbalance, el modelo de ML siempre va a preferir las clases que cuentan mayor número de individuos porque son las que conoce con mayor facilidad, y la predicción para las clases más pequeñas queda casi nula, por esta razón en la **Gráfica 15** se ven las siguientes métricas para Abandono y económica:
Precision = 0 // Recall = 0 // F1-score = 0; son cero porque el modelo no clasificó nada, no le apuntó a nada por la poca ocurrencia que tiene; por lo tanto, **el Resumen Base** es entrenado con el desbalance que hay, qué resultados genera con el desbalance que hay, porque a pesar del desbalance que exista, si se hace un ajuste de parámetros adecuado se puede llegar a algo que sea decente.

G16 *Gráfica 16. Mejor modelo "XGBOOST"*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton



En conclusión **"el base"** significa que el modelo se entrenó con los datos tal cual vienen, sin realizar ningún ajuste al desbalanceo de clases, siendo mejor el modelo el XGBOOST **(Gráfica 16).**

Hoja Resumen hiper: Al mejor modelo XGBOOST se le hizo un ajuste de hiperparámetros **(Gráfica 17)**, pues los hiperparámetros son parámetros que el modelo tiene adentro para hacerlo más o menos profundo, más o menos complejo; entonces, se realizó una sábana de datos, y que él vaya probando diferentes combinaciones de cosas que tiene adentro para que pueda ser un poco más acertado, y como XGBOOST fue el que mejor dio, es al único que se le hizo hiperparámetros, es un proceso costoso cumputacionalmente y se demora, y a pesar de haberse hecho los hiperparámetros, el resultado es el mismo, no mejoró algo y los hiperparámetros no cambiaron mucho respecto a los de base.

G17 *Gráfica 17. Ajuste de hiperparámetros al mejor modelo XGBOOST*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton

VIOLENCIA				
XGBOOST				
index	**precision**	**recall**	**f1-score**	**support**
Abandono	0%	0%	0%	296
Económica	57%	6%	11%	134
Física	37%	31%	34%	6017
Negligencia	41%	47%	44%	4960
Psicológica	34%	6%	10%	4507
Sexual	39%	64%	49%	6761
accuracy			39%	
macro avg	35%	26%	25%	
weighted avg	38%	39%	35%	22675

AGRESOR				
index	**precision**	**recall**	**f1-score**	**support**
Administrativo	0%	0%	0%	55
Escolar	47%	64%	54%	7848
Externos	50%	0%	0%	3140
Familiar	66%	68%	67%	11632
accuracy			57%	
macro avg	41%	33%	30%	
weighted avg	57%	57%	53%	22675

Hoja resumen OVER-UNDER: (Gráfica 18) Lo que se hizo fue manejar el desbalanceo de clases, para lo cual se generó un sobre muestreo **(Técnica Smote)** que lo que hace es generar datos sintéticos o replicar los datos que ya se tienen para alcanzar los valores los valores de las clases más grandes (esto lo hace internamente el algoritmo), y es como si se tuviera un niño con abandono, entonces el algoritmo replica abandono varias veces para llegar a los valores de las clases más grandes, pero esto es dañar los datos, porque cuando el desbalance es tan fuerte, el Smote lo que hace es introducir demasiado ruido al algoritmo, se hizo la prueba con los dos modelos que mejor funcionaron: **XGBOOST y el Perceptor Multicapa.**

G18 *Gráfica 18. Sobremustreo "Técnica Smote"*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton

VIOLENCIA SMOTE

XGBOOST

index	precision	recall	f1-score	support
Abandono	2%	7%	4%	296
Económica	2%	27%	4%	134
Física	38%	20%	26%	6017
Negligencia	37%	57%	45%	4960
Psicológica	31%	21%	25%	4507
Sexual	41%	39%	40%	6761
accuracy			34%	
macro avg	25%	28%	24%	
weighted av	37%	34%	34%	22675

MLP

index	precision	recall	f1-score	support
Abandono	2%	7%	3%	296
Económica	2%	31%	4%	134
Física	36%	17%	23%	6017
Negligencia	38%	57%	46%	4960
Psicológica	29%	27%	28%	4507
Sexual	42%	34%	37%	6761
accuracy			33%	
macro avg	25%	29%	24%	
weighted av	36%	33%	33%	22675

AGRESOR SMOTE

index	precision	recall	f1-score	support
Administrati	2%	20%	3%	55
Escolar	51%	43%	46%	7848
Externos	20%	39%	26%	3140
Familiar	72%	57%	64%	11632
accuracy			50%	
macro avg	36%	40%	35%	
weighted av	57%	50%	52%	22675

index	precision	recall	f1-score	support
Administrati	1%	29%	3%	55
Escolar	52%	36%	42%	7848
Externos	19%	44%	27%	3140
Familiar	72%	55%	62%	11632
accuracy			47%	
macro avg	36%	41%	33%	
weighted av	58%	47%	50%	22675

RESUMEN BASE | RESUMEN HIPER | RESUMEN OVER-UNDER | +

Con el Under **(Gráfica 19)** es al contrario, lo que hace el algoritmo por debajo es un submuestreo, entonces de estos 6017 selecciona 134, de 4960 selecciona 134, de 4507 selecciona 134, y de 6761 selecciona 134 (de cada clase grande selecciona 134); entonces todas las clases quedan muy similares en cantidad, y se vuelve a entrenar el algoritmo aunque se está perdiendo información, porque en este caso se está reduciendo en cada clase 134; en consecuencia con cualquiera de los dos ensayos OVER – UNDER se están dañando los datos de alguna forma.

G19 *Gráfica 19. Submuestreo*

Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton

VIOLENCIA UNDER

XGBOOST

index	precision	recall	f1-score	support
Abandono	2%	15%	3%	296
Económica	2%	40%	3%	134
Física	30%	17%	22%	6017
Negligencia	35%	36%	36%	4960
Psicológica	26%	20%	22%	4507
Sexual	38%	27%	32%	6761
accuracy			25%	
macro avg	22%	26%	20%	
weighted av	32%	25%	28%	22675

MLP

index	precision	recall	f1-score	support
Abandono	2%	17%	3%	296
Económica	1%	43%	3%	134
Física	30%	22%	26%	6017
Negligencia	35%	28%	31%	4960
Psicológica	26%	14%	18%	4507
Sexual	38%	29%	33%	6761
accuracy			24%	
macro avg	22%	25%	19%	
weighted av	32%	24%	27%	22675

AGRESOR UNDER

index	precision	recall	f1-score	support
Administrati	1%	60%	1%	55
Escolar	47%	33%	39%	7848
Externos	16%	29%	20%	3140
Familiar	69%	41%	51%	11632
accuracy			37%	
macro avg	33%	41%	28%	
weighted av	54%	37%	43%	

index	precision	recall	f1-score	support
Administrati	1%	56%	2%	55
Escolar	46%	32%	37%	7848
Externos	15%	33%	21%	3140
Familiar	65%	39%	49%	11632
accuracy			36%	
macro avg	32%	40%	27%	
weighted av	51%	36%	41%	22675

En conclusión, de los experimentos realizados se clasifican dos: XGBOOST con el modelo base (Gráfica 20) y el XGBOOST con el modelo Over-Under (Gráfica 21).

G20 *Gráfica 20. XGBOOST con el Modelo Base*

Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton

VIOLENCIA

LOGISTIC

index	precision	recall	f1-score	support
Abandono	0%	0%	0%	296
Económica	0%	0%	0%	134
Física	37%	33%	35%	6017
Negligencia	40%	47%	43%	4960
Psicológica	34%	5%	9%	4507
Sexual	39%	63%	48%	6761
accuracy			39%	
macro avg	25%	25%	22%	
weighted avg	37%	39%	35%	22675

RANDOM

index	precision	recall	f1-score	support
Abandono	2%	20%	4%	296
Económica	2%	45%	3%	134
Física	37%	14%	20%	6017
Negligencia	38%	45%	41%	4960
Psicológica	29%	23%	26%	4507
Sexual	42%	30%	35%	6761
accuracy			28%	
macro avg	25%	29%	22%	
weighted avg	37%	28%	30%	22675

XGBOOST

index	precision	recall	f1-score	support
Abandonc	0%	0%	0%	296
Económic	57%	6%	11%	134
Física	37%	31%	34%	6017
Negligenc	41%	48%	44%	4960
Psicológic	35%	9%	15%	4507
Sexual	39%	61%	48%	6761
accuracy			39%	
macro avg	35%	26%	25%	
weighted	38%	39%	36%	22675

MLP

index	precision	recall	f1-score	support
Abandonc	0%	0%	0%	296
Económic	50%	4%	7%	134
Física	37%	30%	33%	6017
Negligenc	41%	44%	43%	4960
Psicológic	33%	10%	15%	4507
Sexual	39%	63%	48%	6761
accuracy			39%	
macro avg	33%	25%	24%	
weighted	37%	39%	36%	22675

AGRESOR

index	precision	recall	f1-score	support
Administrativ	0%	0%	0%	55
Escolar	48%	57%	52%	7848
Externos	17%	0%	0%	3140
Familiar	63%	72%	67%	11632
accuracy			57%	
macro avg	32%	32%	30%	
weighted avg	51%	57%	53%	

index	precision	recall	f1-score	support
Administrativ	1%	53%	2%	55
Escolar	50%	38%	43%	7848
Externos	19%	35%	25%	3140
Familiar	73%	51%	60%	11632
accuracy			44%	
macro avg	36%	44%	32%	
weighted avg	57%	44%	49%	22675

index	precision	recall	f1-score	support
Administr	100%	2%	4%	55
Escolar	48%	60%	54%	7848
Externos	44%	1%	3%	3140
Familiar	64%	70%	67%	11632
accuracy			57%	
macro avg	64%	33%	32%	
weighted	56%	57%	53%	22675

index	precision	recall	f1-score	support
Administr	100%	2%	4%	55
Escolar	48%	59%	53%	7848
Externos	37%	2%	4%	3140
Familiar	64%	71%	67%	11632
accuracy			57%	
macro avg	62%	33%	32%	
weighted	55%	57%	53%	22675

RESUMEN BASE | RESUMEN HIPER | RESUMEN OVER-UNDER

G21 *Gráfica 21. XGBOOST con el Modelo Over-Under*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton

VIOLENCIA SMOTE

XGBOOST

index	precision	recall	f1-score	support
Abandono	2%	7%	4%	296
Económica	2%	27%	4%	134
Física	38%	20%	26%	6017
Negligencia	37%	57%	45%	4960
Psicológica	31%	21%	25%	4507
Sexual	41%	39%	40%	6761
accuracy			34%	
macro avg	25%	28%	24%	
weighted av	37%	34%	34%	22675

MLP

index	precision	recall	f1-score	support
Abandono	2%	7%	3%	296
Económica	2%	31%	4%	134
Física	36%	17%	23%	6017
Negligencia	38%	57%	46%	4960
Psicológica	29%	27%	28%	4507
Sexual	42%	34%	37%	6761
accuracy			33%	
macro avg	25%	29%	24%	
weighted av	36%	33%	33%	22675

AGRESOR SMOTE

index	precision	recall	f1-score	support
Administrati	2%	20%	3%	55
Escolar	51%	43%	46%	7848
Externos	20%	39%	26%	3140
Familiar	72%	57%	64%	11632
accuracy			50%	
macro avg	36%	40%	35%	

index	precision	recall	f1-score	support
Administrati	1%	29%	3%	55
Escolar	52%	36%	42%	7848
Externos	19%	44%	27%	3140
Familiar	72%	55%	62%	11632
accuracy			47%	
macro avg	36%	41%	33%	

RESUMEN BASE | RESUMEN HIPER | RESUMEN OVER-UNDER | +

XGBOOST con el Modelo Base por ser el que mejor Accuracy y f1-score tienen a pesar de que no clasifique el abandono (Gráfica 22).

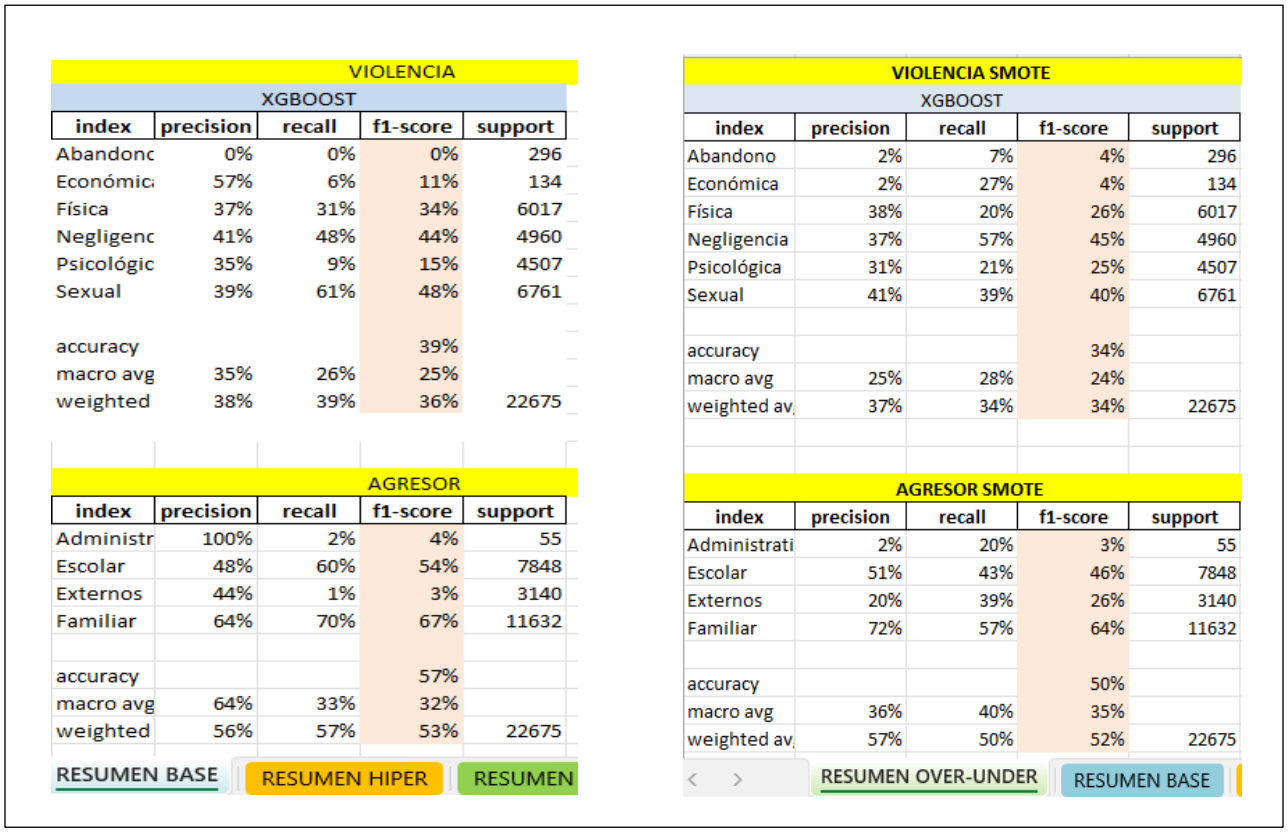
Algunas precisiones:

Precision: mide cuando clasificas algo como positivo, si realmente es positivo. Ejp: si se clasifica a un individuo en clase violencia económica, que realmente haya tenido violencia económica, esto mide la precisión.

Recall: de todos los que estaban en clase abandono económica cuántas detectó.

F1-score: es la media armónica entre ambos. Es como el balance entre los dos, siendo el que resume el resultado de los dos (Precisión y Recall).

G22 *Gráfica 22. XGBOOST con mejor Accuracy y f1-score*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton



VIOLENCIA

XGBOOST

index	precision	recall	f1-score	support
Abandonc	0%	0%	0%	296
Económic	57%	6%	11%	134
Física	37%	31%	34%	6017
Negligenc	41%	48%	44%	4960
Psicológic	35%	9%	15%	4507
Sexual	39%	61%	48%	6761
accuracy			39%	
macro avg	35%	26%	25%	
weighted	38%	39%	36%	22675

AGRESOR

index	precision	recall	f1-score	support
Administr	100%	2%	4%	55
Escolar	48%	60%	54%	7848
Externos	44%	1%	3%	3140
Familiar	64%	70%	67%	11632
accuracy			57%	
macro avg	64%	33%	32%	
weighted	56%	57%	53%	22675

RESUMEN BASE | RESUMEN HIPER | RESUMEN

VIOLENCIA SMOTE

XGBOOST

index	precision	recall	f1-score	support
Abandono	2%	7%	4%	296
Económica	2%	27%	4%	134
Física	38%	20%	26%	6017
Negligencia	37%	57%	45%	4960
Psicológica	31%	21%	25%	4507
Sexual	41%	39%	40%	6761
accuracy			34%	
macro avg	25%	28%	24%	
weighted av	37%	34%	34%	22675

AGRESOR SMOTE

index	precision	recall	f1-score	support
Administrati	2%	20%	3%	55
Escolar	51%	43%	46%	7848
Externos	20%	39%	26%	3140
Familiar	72%	57%	64%	11632
accuracy			50%	
macro avg	36%	40%	35%	
weighted av	57%	50%	52%	22675

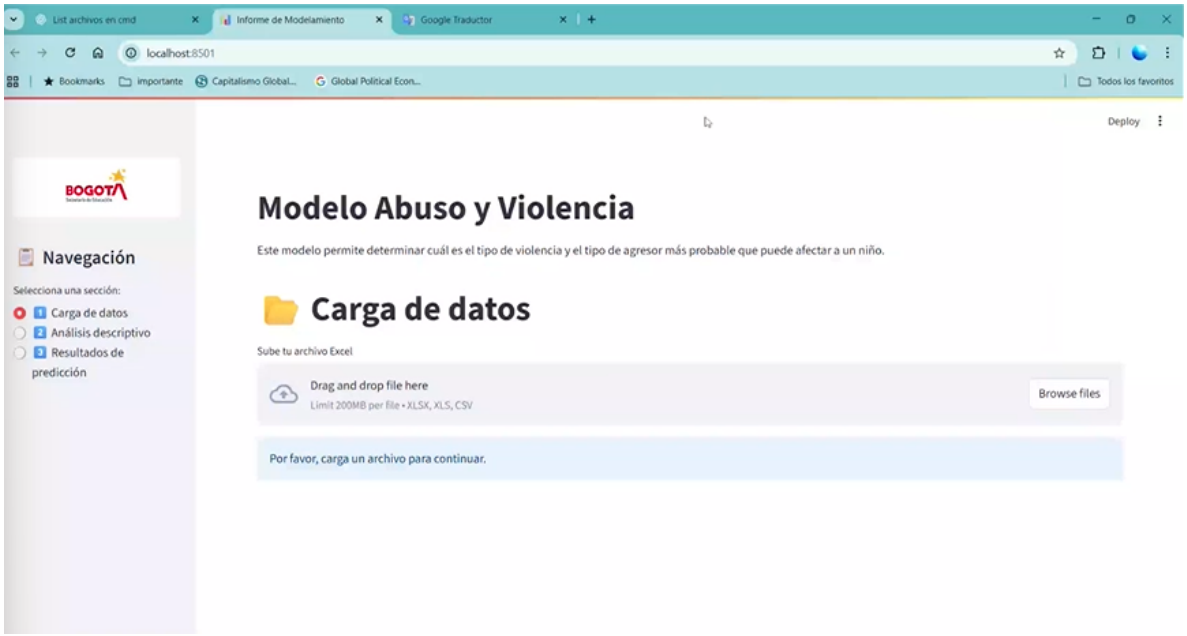
RESUMEN OVER-UNDER | RESUMEN BASE

Diseño final

Como materialización del MockUp se diseñó un tablero sujeto a modificaciones, compuesto por los siguientes **tres módulos:**

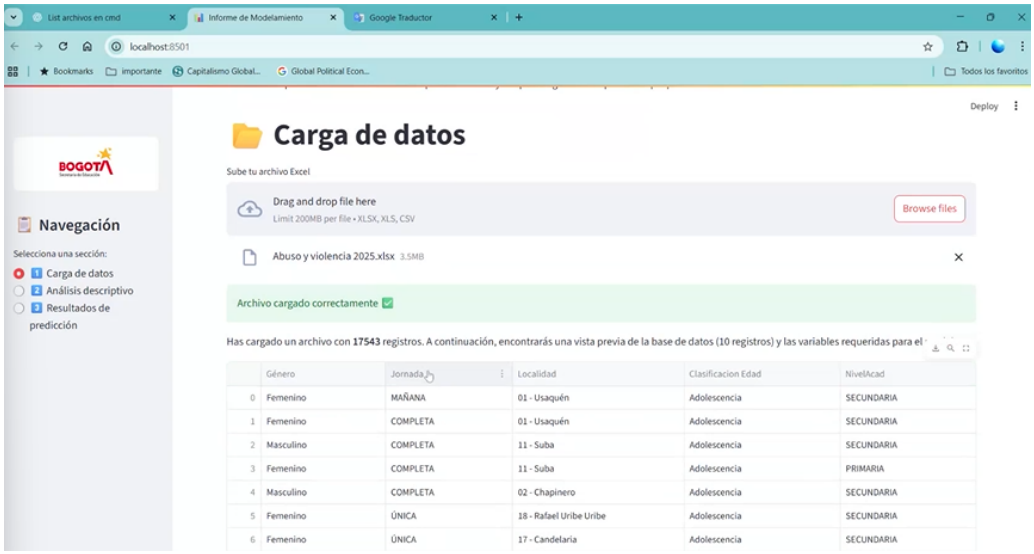
-Carga de datos.
-Análisis descriptivo.
-Resultados de la predicción.

G23 *Gráfica 23. Tablero de predicción para Abuso y Violencia*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton



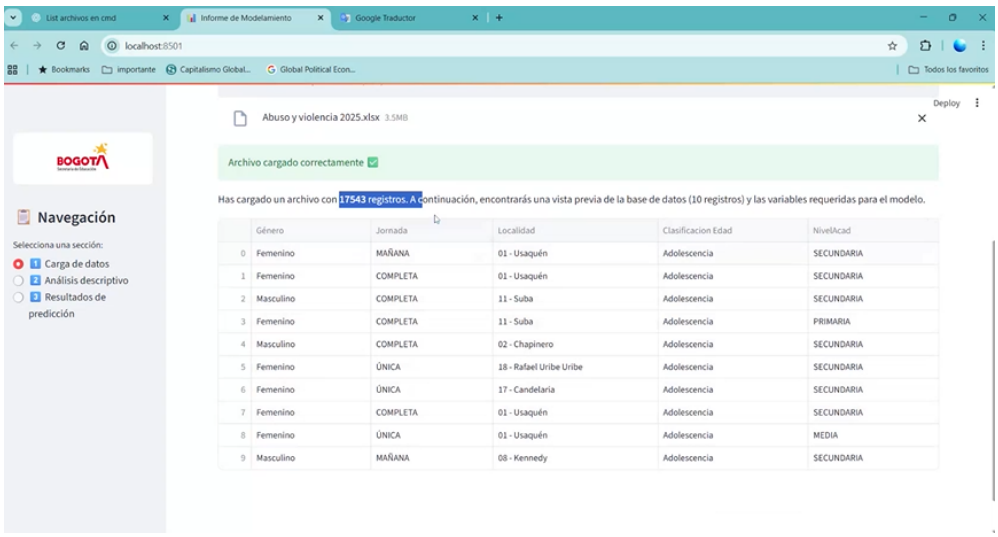
Se cargan los datos de los niños a los cuales se desea hacer la predicción; el archivo debe corresponder a un archivo de Excel, o un CSV o un texto, pero tiene que llevar estas columnas: **Género, Jornada, Localidad, Clasificación Edad, Nivel Académico,** que son las 5 atributos que se están usando para predecir.

G24 *Gráfica 24. Tablero de predicción para Abuso y Violencia – Módulo carga de datos*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton



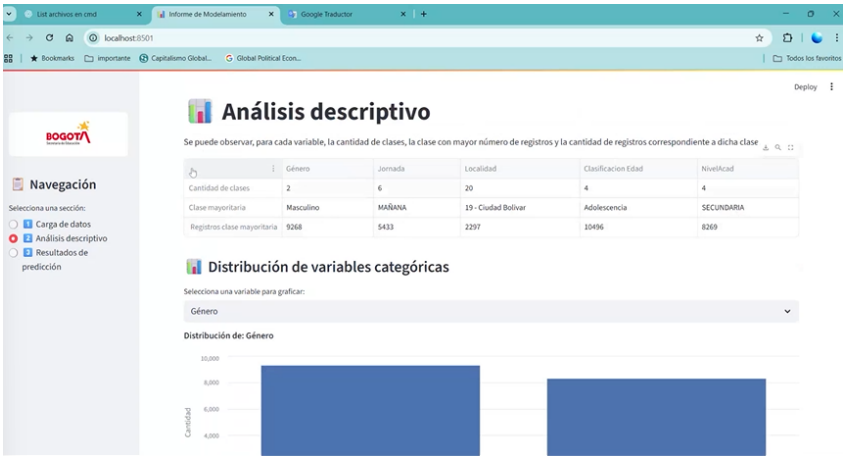
A manera de ejemplo se hizo un cargue con 17543 registros, y se muestra una previsualización de cómo está la tabla cargada.

G25 *Gráfica 25. Módulo carga de datos – Ejemplo datos cargados*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton



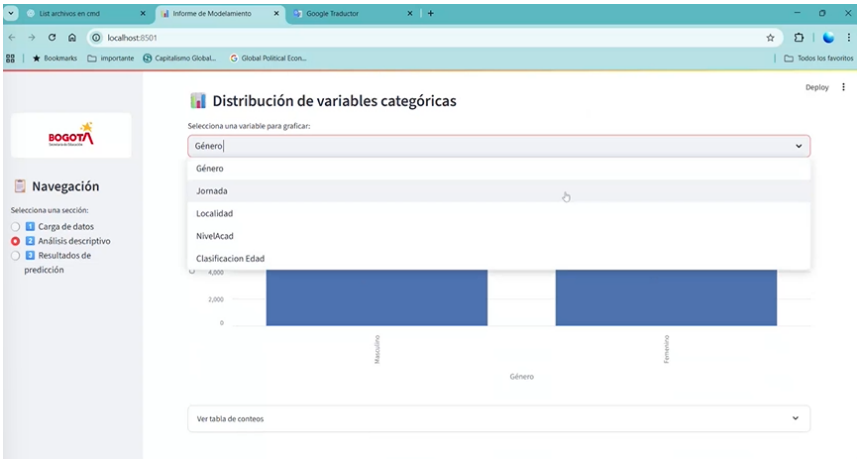
Una vez cargados los datos, se puede ir al módulo de Análisis descriptivo, donde se puede observar un resumen de lo cargado, como: cantidad de individuos por cada atributo; qué clase es mayoritaria y cuántos registros tiene esa clase mayoritaria,

G26 *Gráfica 26. Tablero de predicción para Abuso y Violencia – Módulo análisis descriptivo*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton

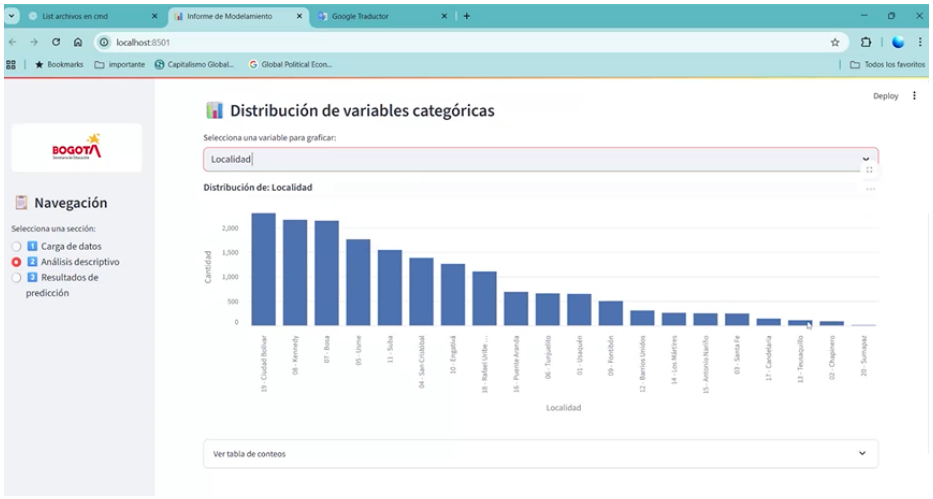


Se puede visualizar por diagrama de barras, seleccionando la variable.

G27 *Gráfica 27. Tablero de predicción para Abuso y Violencia – Distribución de variables categóricas*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton

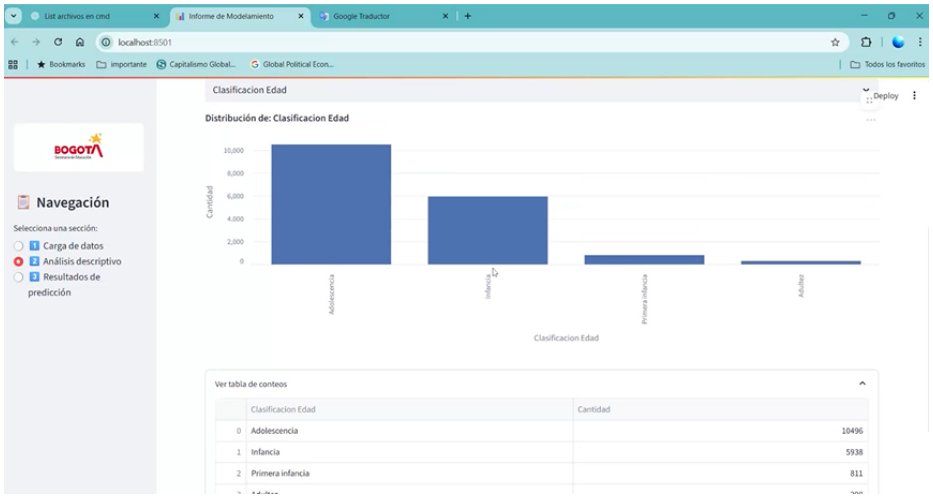


G28 *Gráfica 28. Tablero de predicción para Abuso y Violencia – Distribución por Localidad*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton



Y así, ver cómo se distribuye la población.

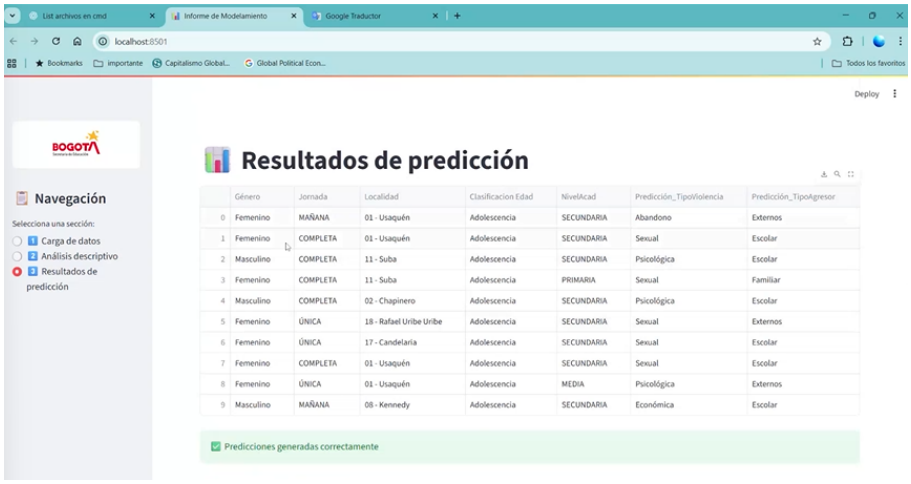
G29 *Gráfica 29. Tablero de predicción para Abuso y Violencia – Tabla de reportes*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton



Y bajo los diagramas de barras se puede observar una tabla con los reportes, la cual se puede exportar.

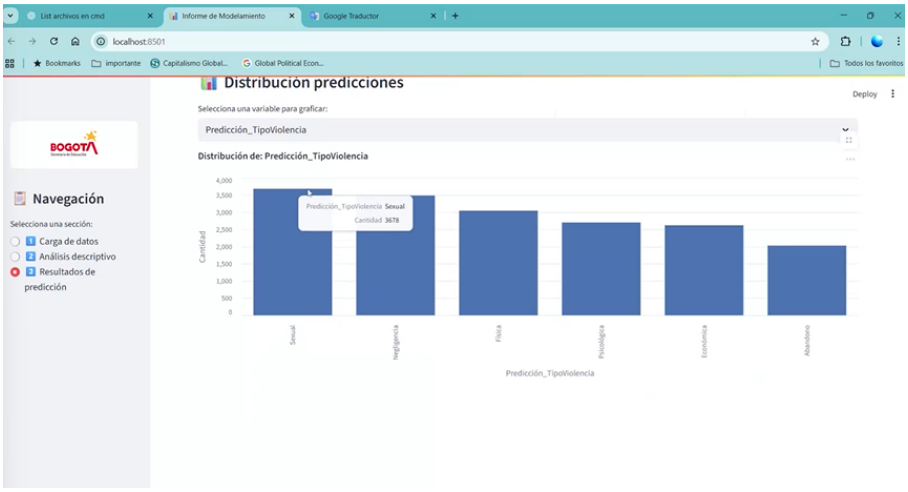
En el siguiente módulo se tienen los resultados de la predicción, con las mismas cinco columnas, y con dos columnas adicionales donde aparece la predicción que hizo el modelo a cada individuo para las variables Violencia y Agresor. **(Gráfica 30).**

G30 *Gráfica 30. Tablero de predicción para Abuso y Violencia – Módulo resultados de predicción*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton

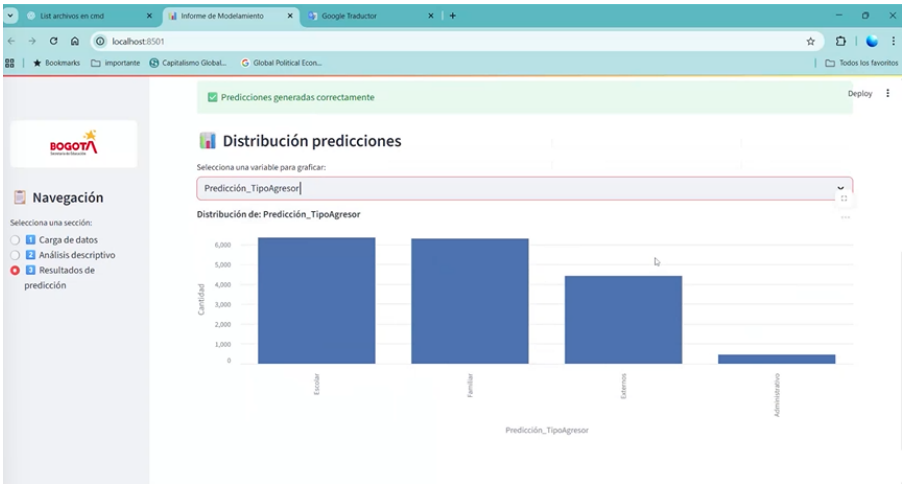


Esta sería la data que se exportaría con las predicciones para cada niño (en un escenario real).

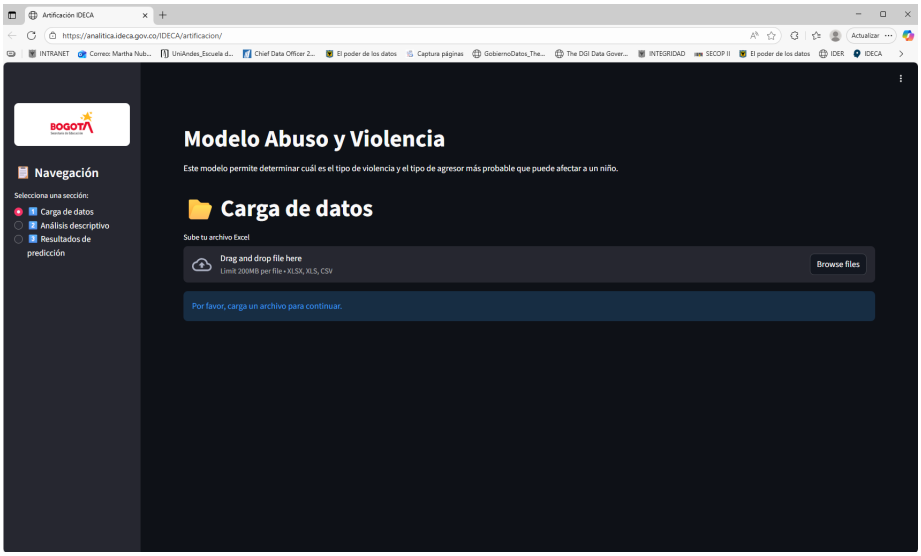
G31 *Gráfica 31. Distribución de la predicción tipo violencia*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton



G32 *Gráfica 32. Distribución de la predicción tipo agresor*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton



G33 *Gráfica 33. Tablero de predicción para Abuso y Violencia – Despliegue en Azure -*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton



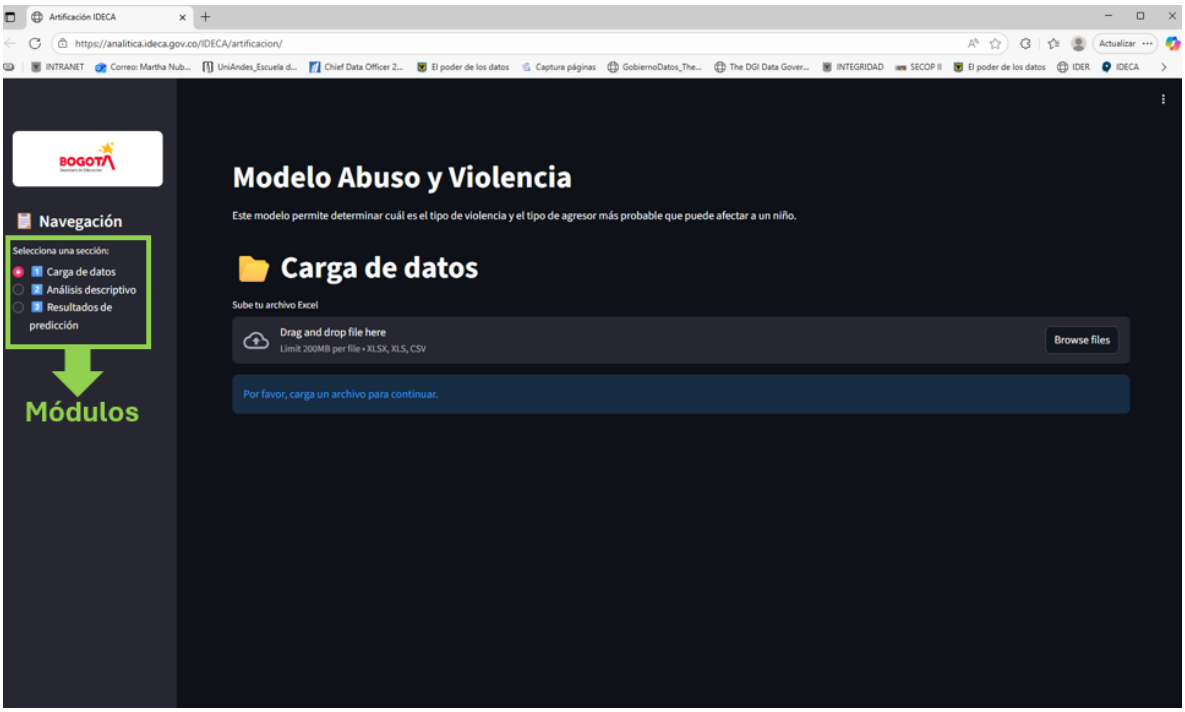
Para el aplicativo se desarrolló el código que soporta el tablero del Proyecto de Artificación con la SED para su despliegue en Azure. Posteriormente, se realizó la creación de un contenedor sobre el servidor Python catavms001, el cual se

configuró con el ambiente Python requerido para la ejecución de la librería streamlit con la que se desarrolló la aplicación.

Se realizó el despliegue de la solución, a la vez que se realizaro las debidas para comprobar su correcta ejecución.

Finalmente, se configuró la URL que aloja la aplicación, quedando de la siguiente forma:
https://analitica.ideca.gov.co/IDECA/artificacion/

G34 *Gráfica 34. Módulos Tablero de predicción para Abuso y Violencia – Despliegue en Azure -*
Fuente: Perfilamiento de Datos y Estadística - Desarrollo Inicial en Phyton



Salvedad
Se están clasificando todos los niños con alguna de las dos violencias, debido a que no se contó con datos de niños que no manifestaran violencia (no hubo data con casos positivos).

Son resultados muy modestos por la falta de información, es decir, los resultados de la predicción son resultados muy genéricos porque si se hubiera tenido información del entorno socio-económico del niño, situación económica de los padres, con quien vive, es decir, poder tener información sobre temas muy puntuales, no solo si tuvo algún tipo de violencia o no, sino del entorno en el que se mueve, se hubiera podido tener mejores resultados, pero con la información a la que se tuvo acceso es a lo único que se pudo llegar; razón por la cual es un buen ejercicio de Incubación con la posibilidad que en el futuro se pueda acelerar y predecir con más y mejores variables.

2. Conclusiones

- De acuerdo con los únicos datos a los que se tuvo acceso, el modelo de ML fue diseñado y desarrollado con los datos que se tienen, por lo tanto, está diseñado para que asigne una categoría de violencia y tipo de agresor a cada estudiante, puesto que no tiene información de estudiantes que no hayan presentado eventos de abuso y violencia, entonces así quedó construido el aplicativo.
- Actualmente, el modelo no pudo contar con información detallada; por lo tanto, para posteriores procesamientos sería ideal y conveniente fortalecer el modelo con información más profunda del estudiante, tal como: nivel de ingresos de los padres o cuidadores, tipo de vivienda, servicio de salud, y lo más importante la coordenada (dirección del estudiante) de manera que se pueda geoprocesar el dato y relacionarlo con otras variables geográficas.

3. Anexo 1: Tipologías Edad Escolar

Outlook

Insumos analítica de datos - Estrategia de artificación

Desde DINA LUZ RIANO CARDENAS <driano@educacionbogota.gov.co>
Fecha Jue 24/07/2025 10:09 AM
Para Martha Nubia Rojas Guerrero <mrojas@catastrobogota.gov.co>; Lillian Daniela Beltran Barrera <ldbeltran@catastrobogota.gov.co>
CC PAULA ANDREA ROMERO ANGARITA <promeroa@educacionbogota.gov.co>; ANDRES LEONARDO URREA VELASQUEZ <alurrea@educacionbogota.gov.co>

1 archivo adjunto (1 MB)
4.-DIR-30-JUN-2025_01072025.xlsx;

Estimadas Martha y Daniela

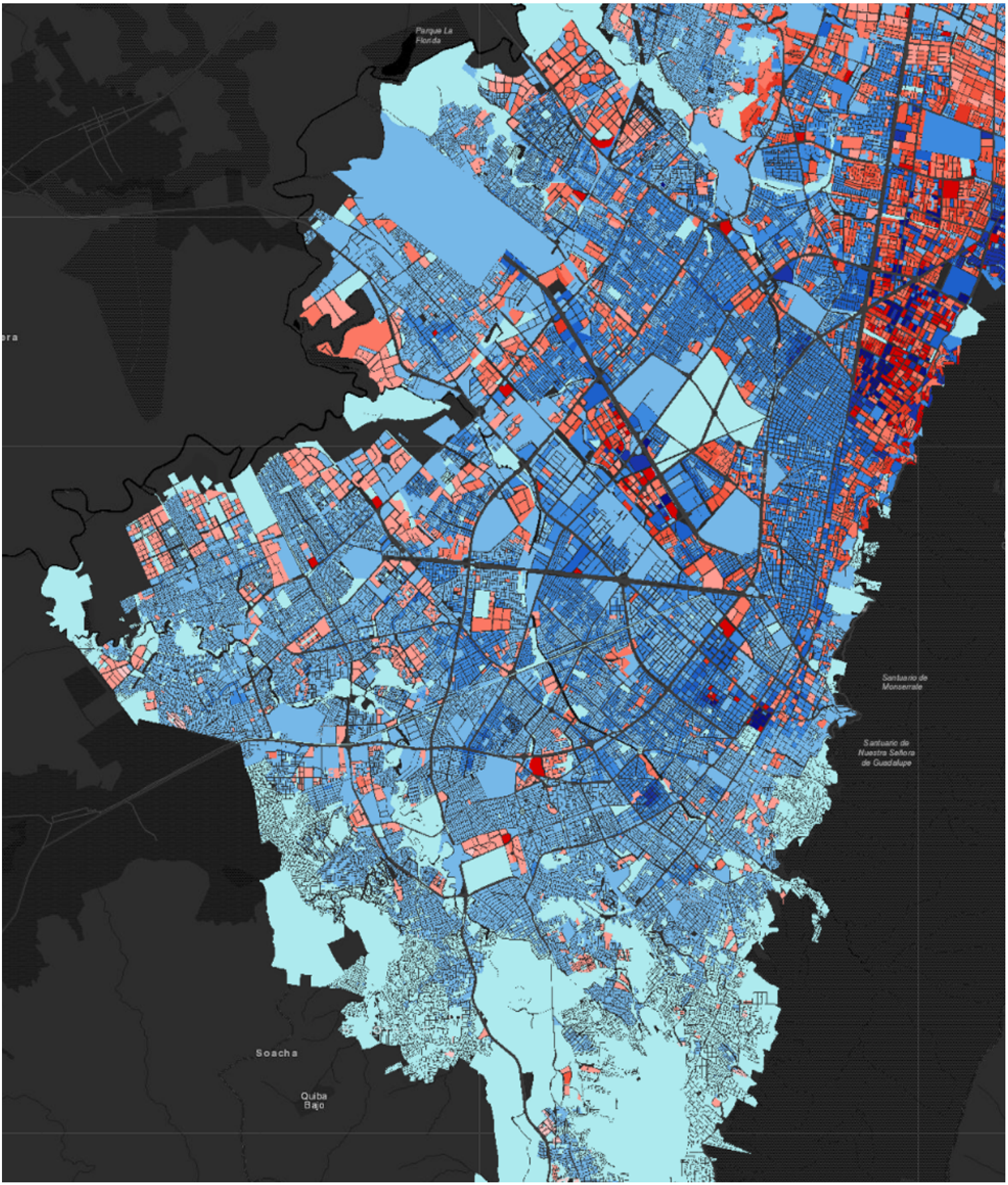
Tal como lo conversamos en la reunión, adjunto:

1. Directorio único de instituciones
2. Rangos de edad que trabajamos desde la Secretaría de Educación Distrital (SED) para el análisis de loa datos
   Primera infancia 0 a 5 años
   Infancia 6 a 11 años
   Adolescencia 12 a 17 años
   Adultez + 18 años
3. Link de Observatorio de Acceso y Permanencia Escolar de la SED
   https://oape.educacionbogota.edu.co/

Seguimos en articulación

Muchas gracias

Por favor evita imprimir, a menos que sea estrictamente necesario, al preferir el uso de documentos digitales se ahorra agua, energía y recursos forestales. La información aquí contenida es para uso exclusivo de la persona o entidad de destino. Está estrictamente prohibida su utilización, copia, descarga, distribución, modificación y/o reproducción total o parcial, sin el permiso expreso de la Secretaría de Educación del Distrito, pues su contenido puede ser de carácter confidencial y/o contener material privilegiado. Si usted recibió esta información por error, por favor contacte en forma inmediata a quien la envió y borre este material de su computador. La Secretaría de Educación del Distrito no es responsable por la información contenida en esta comunicación, el directo responsable es quien la firma o el autor de la misma. www.educacionbogota.edu.co



ideca

ALCALDÍA MAYOR DE BOGOTÁ D.C.
UAECD Catastro Bogotá
BOGOTÁ