

PROTOTIPO DE OBSERVATORIO DE DATOS PARA EL SEGUIMIENTO DEL OBJETIVO DE
EDUCACIÓN DE CALIDAD, CASO DE ESTUDIO: CARACTERÍSTICAS DE LOS
ESTABLECIMIENTOS EDUCATIVOS Y SU INFLUENCIA EN LOS RESULTADOS DE LA
PRUEBA SABER 11.

TRABAJO DE GRADO - PROFUNDIZACIÓN

VIVIAN NATALIA RIVERA SEPÚLVEDA
CARLOS EDUARDO PULIDO PULIDO



PONTIFICIA UNIVERSIDAD JAVERIANA
FACULTAD DE INGENIERÍA
MAESTRÍA EN ANALÍTICA PARA LA INTELIGENCIA DE NEGOCIOS
BOGOTÁ, D.C.
2017

Contenido

RESUMEN.....	4
FASE 1. ENTENDIMIENTO DEL NEGOCIO	5
1. Objetivos del Centro de Excelencia y Big Data CAOBA.....	5
2. Evaluación de la Situación.....	7
2.1. Inventario de recursos.....	7
2.2. Requisitos, supuestos y limitaciones	8
3. Objetivos y Resultados	8
3.1. Objetivos	8
3.2. Resultados esperados	8
4. Manejo de los datos del observatorio	9
5. Entendimiento del objetivo de Calidad de la Educación.	10
FASE 2. ENTENDIMIENTO DE LOS DATOS	14
6. Recopilación Inicial de los Datos	14
7. Descripción de la Base de Datos.....	16
8. Exploración de la Base de Datos.....	18
9. Verificación de la Calidad de los Datos	32
FASE 3. PREPARACIÓN DE LOS DATOS	35
10. Seleccionar la Base de Datos	35
11. Limpieza de la Base de Datos.....	35
11.1. Variables que no aportaban información:	36
11.2. Variables redundantes	36
11.3. Registros inconsistentes e incompletos (imputación)	36
FASE 4. MODELAJE	37
12. Técnicas de modelado seleccionadas.....	37
13. Diseño de prueba de los modelos.....	39
13.1. Prueba Clustering	40
13.2. Prueba Regresión Multinomial	40
14. Construcción del modelo.....	40
14.1. Modelo Clustering	40
14.1. Modelo Regresión Multinomial.....	41
15. Evaluar Modelo	41
15.1. Modelo Clustering	41
15.2. Modelo Regresión Multinomial.....	43

FASE 5. EVALUACIÓN	45
16. Evaluar los resultados	45
17. Revisión del proceso	50
18. Determinar los próximos pasos	51
MANEJO RESPONSABLE DE LA INFORMACIÓN	51
CONCLUSIONES Y RECOMENDACIONES.....	53
BIBLIOGRAFÍA.....	55
ANEXOS.....	55

RESUMEN

Este trabajo aplicado diseña un observatorio de datos que responde al cuarto objetivo de desarrollo sostenible 2015-2030 referente a la Educación de Calidad en Colombia, compuesto de 7 metas y 43 indicadores globales. Esta iniciativa surge por el Centro de Excelencia de *Big Data y Data Analytics* Alianza CAOBA buscando ser pionero en generar una herramienta analítica que genere beneficio social para Colombia. Tomando como marco de referencia el informe de seguimiento a la educación de UNESCO[1] el cual presenta indicadores que responden a una necesidad para América Latina, es necesaria la generación de valor de las fuentes de datos abiertos para la homologación de estos, para el contexto colombiano. Se realiza una exploración inicial de datos con el fin de analizar la relevancia que tienen estas fuentes para el ODS-4, y poder evaluar la calidad con miras a proponer categorización de las variables que definen estos indicadores para Colombia.

El proyecto está basado en la metodología CRISP-DM, y para esto se emplearon técnicas analíticas para la preparación de las fuentes de datos disponibles, asequibles y completas, que permitan el descubrimiento de patrones y relaciones, así como planteamiento de análisis descriptivos e identificación de variables importantes en la educación. Se profundizó en los resultados de las pruebas de saber 11 y se cruzaron con base de los colegios que permitieron, se lograron identificar variables importantes que tienen influencia en los puntajes de las competencias en lectura crítica y matemáticas. Luego generar regresiones multinomiales por periodo, competencias y departamentos, se llegó a la conclusión que el nivel de formación de los docentes de los colegios tienen efecto en los puntajes, los colegios con mejores resultados tiene una marcada diferencia en la segunda lengua, el nivel educativo de las madres, específicamente posgrado aumenta la probabilidad de que un estudiante tenga desempeño alto y finalmente, la edad del estudiante en el momento de presentar el examen genera mejores puntajes

Para la estructuración del observatorio, se realizó una integración de herramientas especializadas en la publicación de datos abiertos, enmarcado en un esquema de gobierno de datos y un listado de fuentes que respondan a cada meta, en paralelo se aprovecharon recursos tecnológicos de software para el diseño e implementación de la integración de la información, para finalmente generar una prueba de concepto del observatorio de datos.

FASE 1. ENTENDIMIENTO DEL NEGOCIO

1. Objetivos del Centro de Excelencia y Big Data CAOBA

En el año 2015 se firmó la agenda 2030 para el Desarrollo Sostenible por la Asamblea General de Naciones Unidas conformada por 193 países, entre los que se encuentra Colombia. Tal como la define la UNESCO, “esta hoja de ruta es la oportunidad histórica para América Latina y el Caribe, ya que incluye temas altamente prioritarios para la región”[1] constituida por 17 objetivos de desarrollo sostenible –ODS- y 169 metas, sin embargo será el objetivo número cuatro (Educación de Calidad), el marco conceptual de este trabajo aplicado. La agenda anteriormente mencionada, describe el cuarto objetivo como:

“Garantizar una educación inclusiva, equitativa y de calidad y promover oportunidades de aprendizaje durante toda la vida para todos”

Menciona también que la consecución de una educación de calidad es la base para mejorar la vida de las personas y el desarrollo sostenible y aunque algunos países han producido importantes avances con relación a la mejora en el acceso a la educación y también se ha incrementado el nivel mínimo de alfabetización, es importante *redoblar* los esfuerzos para conseguir mayores avances en este objetivo universal. En este sentido, en el año 2016 el instituto de estadística de la UNESCO basado en el ODS –4, publicó un importante informe de seguimiento de educación en el mundo, el cual revela el enorme desfase que existe entre la situación actual del mundo en materia de educación y lo que este se ha comprometido a lograr de aquí a 2030.

Tal como lo menciona el informe de la UNESCO, Colombia es país pionero en la aplicación de los ODS, el gobierno estableció una comisión interinstitucional para preparar e implementar la agenda de desarrollo de estos, esta comisión refleja el fuerte compromiso político para la aplicación intersectorial de la agenda. A los gobernadores y alcaldes se les alentó para que los incluyeran en sus planes de desarrollo y específicamente se planteó en el Plan Nacional de Desarrollo convertir a Colombia en el país más educado de América Latina en el año 2025. Lo anterior, implica que existe en el país gran interés sobre el cumplimiento de los ODS y como lo menciona la Secretaría Ejecutiva de CEPAL, “esta agenda al ser ambiciosa y visionaria, requiere de la participación de todos los sectores de la sociedad y del Estado para su implementación”. [1]

Por otro lado, en el año 2016 se creó Alianza CAOBA, un Centro de Excelencia y Apropiación de Big Data y Data Analytics, conformado por entidades públicas, tales como DNP y privadas para promover el Big Data y la analítica en Colombia, esencialmente este centro combina la capacidad de investigación de las universidades que la conforman (ICESI, EAFIT, Andes y Javeriana), la oferta de productos líderes mundiales en tecnología de las empresas que también hacen parte (IBM Colombia, SAS Colombia, EMC y Clúster CREATIC) con el objetivo de generar servicios y soluciones innovadoras que agreguen valor a los sectores estratégicos del país.

De acuerdo con lo anterior y entendiendo el quehacer de la Alianza CAOBA y las relaciones que la fundamentan, es una organización idónea para plantearse ideas analíticas que permitan evaluar la

calidad, disponibilidad, completitud y pertinencia de las fuentes de información de educación con qué cuenta hoy el país. Sus actividades se concentran en cuatro ejes:

1. Investigación Aplicada fundamentada en problemas relevantes para el país y desarrollada con la rigurosidad científica requerida para trascender en el área de conocimiento y posibilitar la transferencia tecnológica de las soluciones creadas.
2. Consultoría que favorezca la apropiación de tecnologías, metodologías y herramientas de análisis y procesamiento de grandes volúmenes de información en el país y la región.
3. Formación que integre los saberes y experiencia de las partes que la conforman para contribuir de manera efectiva en la apropiación de las temáticas asociadas a Big Data y Data Analytics en el país.
4. Apoyo de iniciativas de emprendimiento y spin-offs colombianos cuya propuesta de valor esté fundamentada en la generación de soluciones alrededor de las temáticas concernientes a CAOBA.

A partir de la política de datos abiertos adoptada por Colombia y en la cual según la Organización para la Cooperación y el Desarrollo Económicos (OCDE) que publicó los resultados del OURData Index, en el que Colombia ocupó el cuarto lugar después de Corea, Francia y Japón. El país superó en el listado a naciones como Reino Unido, Estados Unidos, Alemania y España, gracias al trabajo que ha venido realizando con los datos abiertos provenientes del sector público, en el marco de la estrategia enfocada hacia un Gobierno abierto y más transparente.

De acuerdo con lo anterior se identifica que existe una necesidad de explorar las fuentes de datos disponibles que permitan responder a los ODS, tal como se mencionó anteriormente, la UNESCO identificó que existe a nivel global un desfase entre las estadísticas actuales de educación y lo que se espera lograr en 2030, pero se desconoce la amplitud de la brecha que tiene actualmente y específicamente Colombia para responder a dichas metas. Con esta motivación Alianza CAOBA se ha propuesto ser pionera en la exploración de dicha información que permita diagnosticar el avance de Colombia en este tema, su primera propuesta, como Centro de Excelencia, es el diseño de una herramienta que integre y permita visualizar la información pertinente de Educación en Colombia, haciendo un aporte a la solución de problemáticas relevantes para el país, lo cual favorecerá la visibilidad del Centro de Excelencia a nivel nacional e internacional.

Ahora bien, dentro del punto de vista analítico, se desconoce la utilidad de estas fuentes de información para generar análisis relevantes para el cumplimiento de los indicadores de Educación de Calidad expuesto en los ODS, por tal razón con este estudio se pretende responder; *con base en las fuentes de información disponibles y abiertas, ¿cómo se puede tener un acercamiento a los indicadores de la educación de calidad en Colombia enmarcados en las metas del ODS-4 que generen valor a dichas fuentes?*

El proyecto planteado tal como se describe anteriormente permitirá afrontar una situación real de país, donde se deben identificar las fuentes de datos necesarias para desarrollar el problema, evaluar la calidad de las mismas y generar aprovechamiento de los registros administrativos en el país aplicando un modelo analítico. Es importante aclarar que este proyecto profundizará en los análisis de valor agregado para *una* meta conforme a las fuentes disponibles.

2. Evaluación de la Situación

Para el desarrollo del proyecto se requieren dos perfiles particulares, uno que tenga conocimiento sobre el manejo de herramientas especializadas para la construcción de una plataforma que permita la visualización y presentación de resultados de interés de Educación de Calidad y, por otro lado, un perfil con conocimientos y experiencia en educación nacional, indicadores de educación y fuentes existentes pertinentes para análisis de educación en todos los niveles de escolaridad, desde preescolar hasta secundaria y mercado laboral. Dado que el recurso temático con el cual se cuenta inicialmente, es la lista de metas e indicadores propuestos por la UNESCO y el informe de seguimiento a la Educación en el mundo realizado por la misma organización el cual se debe considerar para realizar la caracterización de los indicadores Colombia.

2.1. Inventario de recursos

Participantes:

Carlos Pulido, Ingeniero de Sistemas de la Universidad Central y especialista en sistemas gerenciales de ingeniería de la Universidad Javeriana, con experiencia en la generación de información para reportar la información al Ministerio de Educación en la Universidad Javeriana, y actualmente ingeniero de proyectos en la misma institución dedicado al mantenimiento y desarrollo de aplicación para toda la Universidad, actualmente encargado de soportar el sistema de Business Intelligence de la Universidad en cuanto a incidentes y algunos desarrollos nuevos, desde la Dirección de Tecnologías de Información de la Vicerrectoría Académica.

Natalia Rivera, Profesional Estadística de la Universidad Santo Tomás, experiencia como estadística en la Coordinación de Censos y Demografía del Departamento Administrativo Nacional de Estadística específicamente en el procesamiento de entradas y salidas del país de Migración Colombia para la estimación flujos migratorios, en la Oficina de Autoevaluación y Acreditación de la Universidad Antonio Nariño desarrollando análisis estadísticos para la presentación de programas en renovación de registro calificado y acreditación institucional y actualmente trabaja en la Secretaría de Planeación – Rectoría de la Universidad Javeriana como profesional para el desarrollo de estudios institucionales.

Tecnologías:



1



rapidminer

2



Microsoft®
SQL Server



Fuentes: Bases de datos abiertos, relacionadas con educación que permitan responder con los indicadores del Objetivo de Calidad, o fuentes oficiales que permitan tener un acercamiento a análisis de valor agregado para los indicadores. Las fuentes se relacionarán en el anexo 1.

2.2. Requisitos, supuestos y limitaciones

Aunque no hay tramites de confidencialidad, las fuentes de información son las disponibles de datos abiertos. Para esta fase del proyecto no se solicitará información adicional a la disponible en la web, sin embargo, la caracterización de indicadores identificará las fuentes necesarias para su cálculo, para las cuales se recomienda que el Centro de Excelencia CAOBA desarrolle estrategias para solicitud de la información.

Existe una limitación de tiempo para la ejecución del proyecto, en este sentido, se profundizará en una meta conforme a la información disponible, a la cual se le aplicará desde la segunda fase de la metodología CRISP-DM.

3. Objetivos y Resultados

3.1. Objetivos

General

Diseñar un observatorio de datos de educación en Colombia, a partir de la información disponible y abierta que responda al cuarto objetivo de desarrollo sostenible 2015-2030.

Específicos

- Identificar las fuentes de información disponibles para categorizar las variables que definen las metas propuestas en el 4 objetivo de desarrollo sostenible para el contexto colombiano.
- Definir la política de gobierno de datos que permita gestionar los datos del observatorio
- Plantear análisis *Ad hoc*, con base en la información recopilada, que permitan un acercamiento a los indicadores de la educación de calidad en Colombia.
- Desarrollar una prueba de concepto del observatorio para la visualización de información de educación de calidad.

3.2. Resultados esperados

A continuación, se relacionan los productos esperados para cada etapa de la metodología CRISP-DM y el respectivo objetivo específico.

Tabla 1. Resultados y productos esperados

Metodología	Resultados Esperado
Comprensión del negocio/ objetivo específico 1.	Caracterización de los indicadores que propone la UNESCO para las 7 metas del cuarto objetivo de desarrollo sostenible
Comprensión de los datos/ objetivo específico 2.	<ul style="list-style-type: none"> • Diseño de la política de gobierno de datos • Análisis descriptivos de las fuentes de datos.
Preparación de los datos	Vista minable de las fuentes de información para la meta seleccionada de acuerdo con la completitud del anterior paso.
Modelado/ objetivo específico 3.	<ul style="list-style-type: none"> • Planteamiento y desarrollo de los análisis <i>ad hoc</i>. • Diseño del observatorio de datos
Evaluación	Evaluación de los modelos planteados identificando el aporte a los indicadores de educación de calidad planteados por la UNESCO.
Despliegue/ objetivo específico 4.	<ul style="list-style-type: none"> • Prueba de concepto del observatorio de datos, visualización de información para una meta del ODS-4. • Recomendaciones finales para la completitud del observatorio

Fuente. Elaboración propia

4. Manejo de los datos del observatorio

La iniciativa que tiene el centro de Excelencia COBA, es un esfuerzo de largo aliento orientado a generar una herramienta que permita visualizar información relevante relacionada a los ODS, y se plantea etapas para lograr este gran objetivo del negocio, puntualmente, la etapa inicial es este trabajo aplicado, que pretende realizar una fundamentación y hacer definiciones generales que den base al resultado general, un observatorio de datos para los ODS aplicados a Colombia.

Dadas las características de la necesidad del negocio, se hace visible la importancia que tiene los datos en el proyecto, y dado que estos son la fuente primaria para este proyecto, se hace necesario identificar mecanismos que permitan la interacción de los datos entre: los análisis específicos que se puedan generar, la visualización de los datos en el Observatorio de datos y el consumo de estos datos desde fuentes externas, para la información aplicable a ser compartida, con la aplicación de estrategias que permitan una orquestación idónea entre los diferentes actores.

Con esta necesidad clara del negocio, se plantea como uno de los entregables el documento de gobierno de datos, **Producto CAOBA No 1. Política de Gobierno de datos**, que dará las bases a todo el manejo de datos para el observatorio de datos que es el propósito central del Centro de Excelencia Caoba.

Este documento de gobierno de datos, permitirá tener lineamientos claros de cómo se deben organizar y estructurar la información, así como las políticas que la organización debe adoptar, permitiendo que el proyecto actual y los próximos, que abordarán alguno de los otros objetivos de desarrollo sostenible,

tengan herramientas claras de cómo articularse a los datos existentes y poderle dar continuidad al proyecto de la manera más transparente, complementando el observatorio.

Este manejo de datos se complementará con la experiencia obtenida en la instalación, configuración y puesta en funcionamiento de la prueba piloto, que aplicará toda la política de gobierno de datos validando este documento y permitiendo se identifiquen, desde la práctica, puntos de mejora y ajustes, cumpliendo así con una de las políticas de este documento enfocada a la revisión continua y periódica que se debe tener.

5. Entendimiento del objetivo de Calidad de la Educación.

Cómo se ha mencionado anteriormente, este proyecto tiene como objetivo específico caracterizar los indicadores y metas que componen el cuarto objetivo de educación de calidad, es decir, describir los conceptos incluidos de forma explícita o implícita en la formulación de la meta y la articulación de estos conceptos en las formulación de los indicadores propuestos (mundiales y temáticos), relacionar la metodología de cálculo, así como la identificación de fuentes y herramientas disponibles en la actualidad para recopilar datos relevantes y de esta manera, proponer un análisis que genere valor a los resultados de los indicadores y las fuentes oficiales disponibles.

El cuarto objetivo de desarrollo sostenible pretende: Garantizar una educación inclusiva, equitativa y de calidad y promover oportunidades de aprendizaje durante toda la vida para todos, contiene 7 metas (ver ilustración 1) y 43 indicadores.

Ilustración 1. Metas del objetivo de Educación de Calidad



Para el Centro de Excelencia CAOBA, es relevante ahondar en cada uno de los indicadores para lo cual se desarrolló un documento de caracterización de indicadores **Producto CAOBA No. 2**, con base en el documento de la UNESCO[2], el cual presenta la siguiente estructura:

- **Definición Metas**, por ejemplo: Meta 1. Para 2030, velar porque todas las niñas y todos los niños terminen los ciclos de la enseñanza primaria y secundaria, que ha de ser gratuita, equitativa y de calidad y producir resultados escolares pertinentes y eficaces.
- **Definición del indicador**, se relacionan los indicadores de cada meta, mundiales y temáticos. Por ejemplo: Porcentaje de niños y de jóvenes: a) en los cursos segundo/tercero; b) al final de la enseñanza primaria; y c) al final del primer ciclo de la enseñanza secundaria, que alcanzan al menos un nivel mínimo de dominio de: i) la lectura y ii) las matemáticas, por sexos.
- **Contexto y limitaciones del indicador**, cada indicador contiene terminología específica que según el caso se debe homologar para Colombia o incluso el indicador presenta limitaciones de cálculo a nivel mundial los cuales se deben reconocer. Por ejemplo, actualmente no existen estándares comunes válidos por la comunidad internacional para evaluar dominio mínimo de lectura y matemáticas.
- **Propósito del indicador**, por ejemplo: El indicador es una medida directa de los resultados de aprendizaje alcanzados en las áreas evaluadas en las etapas pertinentes de la educación.
- **Método de cálculo del indicador**, presenta la fórmula y variable que compone cada indicador. Por ejemplo,

$$\text{Rendimiento por encima del nivel mínimo, } PL_{tn,s\text{mínimo alcanzado}} = p$$

Donde,

- p** Porcentaje de estudiantes en una evaluación de aprendizaje.
- n** Etapa de educación, a, b y c.
- s** Lectura o matemáticas
- t-i** El año de análisis t, donde i este entre 0 y 5.
- Smin** Desempeño mínimo en las pruebas

- **Desagregación propuesta**, a esta desagregación se propone tener el indicador. Por ejemplo, Por edad o grupo de edad de los estudiantes, sexo, ubicación, situación socioeconómica, estatus migratorio y origen étnico. La situación de la discapacidad no está actualmente disponible en la mayoría de las evaluaciones de aprendizaje nacionales e internacionales.
- **Fuentes recomendadas por la UNESCO**, desde el conocimiento de la UNESCO se proponen fuentes que se pueden utilizar para responder el indicador, por ejemplo: resultados de las pruebas TERCER.
- **Fuentes disponibles**, dentro de la identificación de la fuente que responda al indicador, es posible encontrar casos donde la fuente propuesta por la UNESCO no aplica para Colombia, este parte de realizará una exploración de las fuentes disponibles.

Tabla 2. Relación de fuentes para el indicador 1 de la Meta 1.

Fuente	¿Qué es?	Disponibilidad	Formato	Observaciones
U - TERCE (Tercer Estudio Regional Comparativo o Explicativo) De: UNESCO	El objetivo principal es obtener datos comparativos sobre el desempeño escolar de los estudiantes de grados tercero y sexto de América Latina y el Caribe en las áreas de Ciencias, Matemáticas y Lectura.	Bases de datos publicadas en la página web de la UNESCO. http://www.unesco.org/new/es/santiago/terce/databases/ Bases de datos en el FTP del ICFES. ftp://ftp.icfes.gov.co/SABER1/1/SB11-BASES%20DE%20DATOS/	SPSS, STATA, TXT	Se cuenta con la información del año 2013. Para ingresar al FTP se necesita: Usuario: ftpdbicfes Contraseña: dataicfes\$%

Este resultado permite dimensionar el alcance del objetivo de Educación de Calidad, así como las oportunidades de mejora de las mediciones propuestas. Dentro de la primera revisión a los indicadores se logró identificar que los Objetivos de Desarrollo Sostenible tienen gran interés en el análisis que permita la comparación de países a nivel mundial o al menos Latinoamérica, y esto limita la inclusión de fuentes colombianas disponibles para responder sobre la Calidad de la Educación.

En este sentido, tal como se mencionó en el planteamiento del proyecto, se realizará un análisis *ad hoc* sobre educación en Colombia que permita generar valor agregado a las fuentes disponibles, así como aportar mayor información al estado actual de Colombia en algunas de las metas de educación de calidad en Colombia. Con base en la revisión primaria que se ha realizado a las fuentes de datos que existen para responder a cada indicador, es posible identificar que, en algunos casos, Colombia no cuenta actualmente con información tal como lo propone la UNESCO, en este sentido, se identificaron cuatro metas en las cuales con fuentes colombianas es posible plantear análisis que permitan un acercamiento a estado actual de Colombia en estos temas. Ver ilustración 2.

Finalmente, luego de presentar las propuestas a la Organización, se concluyó que no existe diferencia, ni preferencia para ellos para la selección de alguno de los análisis, en este ejercicio es importante presentar acercamiento de análisis de educación en Colombia relacionados con las metas de los ODS y realizar un despliegue estratégico en un prototipo funcional de Observatorio de Datos, siendo así, se trabajará con la meta de alfabetización y conocimiento elementarles de aritmética, más adelante se profundizará en las hipótesis planteadas en esta temática.

Ilustración 2. Propuestas de análisis a partir de la información disponible



Meta: Alfabetización y conocimiento elementales de aritmética

- Cuáles variables influyen en los resultados de las pruebas 11, por departamento, estrato, clasificación colegio.
- Fuente saber 11, resultados microdato 17 años, apróx 10.000.000 de registros
- Características del colegio, casa, hogar, discapacidades estudiantes, sociodemográficos, trabajo, veces de presentar el examen, nivel educativos de la madre, padre, ingresos, puntajes.
- Reducción de dimensionalidad, clustering y regresión multinomial.



Meta: Competencias para el trabajo

- Influencia de la competencias TIC de los egresados de pregrado en los ingresos después de año de grado.
- Fuente encuesta de seguimiento a egresados , 15.450 registros(factor de expansión), aplicada en 2014
- Información personal, familiar, limitaciones, idiomas, plan de vida, movilidad interna o externa, situación laboral, nivel de identidad con la institución
- Regresión lineal



Meta: Enseñanza técnica, profesional superior y de adultos

- Valor agregado de la instituciones de educación superior dependiendo del colegio de procedencia.
- Fuente Resultados Saber 11 (02-11) y Saber PRO (11-14).
- Base de cruce de estudiantes aprox 220.000
- Reducción de dimensionalidad y regresión múltiple.



Meta: Alfabetización de adultos

- Influencia del nivel educativo de los padres en los resultados de las pruebas 11.
- Fuente saber 11, resultados microdato 17 años, apróx 10.000.000 de registros
- Características del colegio, casa, hogar, discapacidades estudiantes, sociodemográficos, trabajo, veces de presentar el examen, nivel educativos de la madre, padre, ingresos, puntajes.
- Reducción de dimensionalidad, emparejamiento y regresión múltiple.

FASE 2. ENTENDIMIENTO DE LOS DATOS

El objetivo de esta fase es reunir datos iniciales de la información, familiarizarse con los datos e identificar la calidad de los datos, así como detectar subconjuntos interesantes para formar algunas hipótesis. En este sentido, desde esta fase del proyecto se realizarán acercamientos analíticos relacionados cuáles son las variables que influyen en los resultados de las pruebas saber 11 (Meta 6. Alfabetización y conocimiento básicos en aritmética), para lo que se utilizará principalmente los resultados de las pruebas saber 11, denominado anteriormente ICFES.

La prueba de saber 11 se realiza a toda la población, no es necesario utilizar ponderaciones para estimaciones poblaciones y en este sentido se reducen errores estadísticos, permite generar interpretaciones a mayor desagregación que la nacional, incluso permite llegar a nivel de urbano/rural de los municipios colombianos y finalmente, dado que es una prueba importante para los estudiantes siendo utilizada para la admisión de la Instituciones de Educación Superior esto hace que los estudiantes se esfuercen para obtener los mejores resultados y no se convierta en un requisito de grado, como sucede actualmente en las pruebas Saber PRO.

6. Recopilación Inicial de los Datos

Los datos que se utilizarán corresponden a los resultados del examen saber 11, desde el año 2.000 primer semestre y hasta el año 2.017 primer semestre. El ICFES dispone de una plataforma pública en la página web³ para los investigadores con el acceso a los microdatos de las pruebas de estado. La información se encuentra disponible en archivos comprimidos (.zip) que contienen archivos de texto planos con separadores, el separador utilizado no es el mismo en todos los archivos; cada archivo contiene la información recolectada por semestre, de tal forma que serán 35 archivos para ser consolidados.

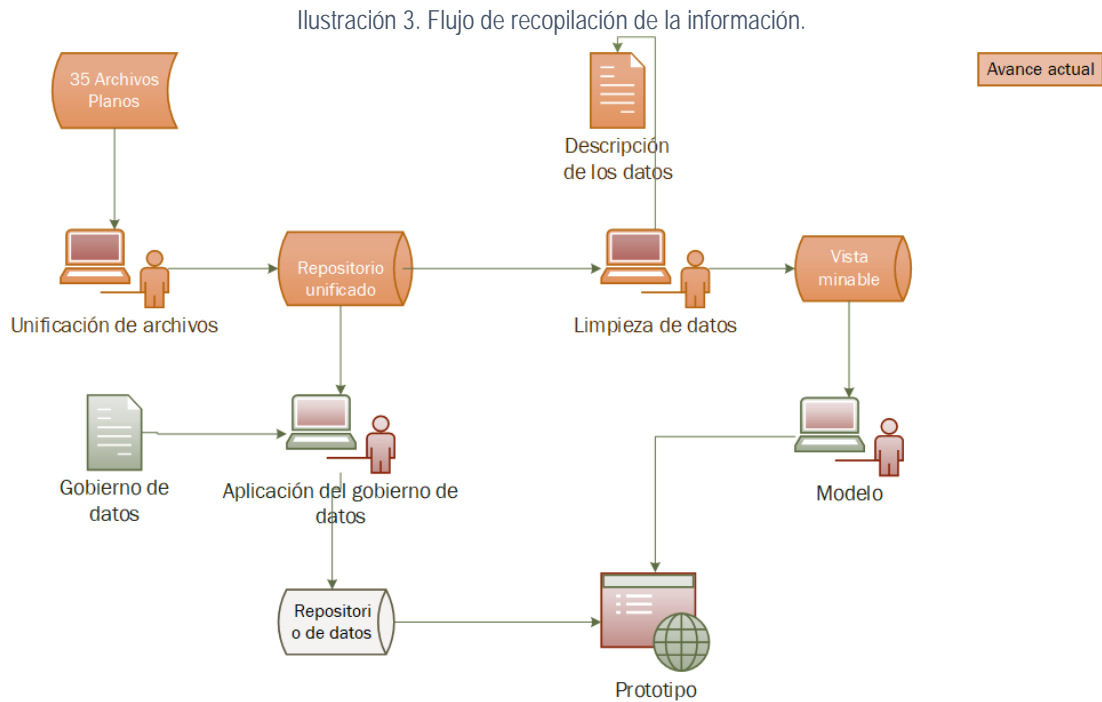
La estructura de cada archivo va cambiando a medida que cambia la fecha de registro de la información, es decir que con el paso del tiempo el número de columnas del archivo aumenta y en algunos casos sus nombres cambian, de tal forma que en años diferente se puede llamar a un mismo campo de diferentes maneras. Se identifica que para los periodos entre 2000 primer semestre y 2004 segundo semestre se mantiene una continuidad en el número de variables por archivo (28 en promedio), luego se identifica otro grupo de archivos con 54 variables en promedio entre el 2005 segundo semestre y el 2008 del mismo semestre, los siguiente tres periodos cuentan con 83, 100 y 88 variables respectivamente entre el periodo 2009 primer semestre y el 2010 primer semestre, continuando de menor a mayor se identifican 5 periodos con 96 variables, luego 3 con 69 y finalmente un grupo de archivos con variables entre 71 y 95 para un promedio de 83. Luego de la consolidación inicial se identificaron 234 variables y 9.285.527 registros.

³ <http://www.icfes.gov.co/investigadores-y-estudiantes-posgrado/acceso-a-bases-de-datos>

Los archivos comprimidos ocuparon 567.116.694 bytes y al descomprimirlos ocupan 3.400.224.685 bytes en disco; en promedio los archivos tienen un tamaño 98.433.685 bytes, el detalle de los tamaños por archivo se encuentra en el **Anexo 2**.

Luego de realizar labores de procesamiento y cargarlas en una base de datos SQL Server 2016 de Microsoft ocupó un tamaño cercano a los 7 GB, este incremento de tamaño se debe al número de variables (columnas) de la tabla que unifica toda la información. La codificación de los archivos está en UTF-8 por lo que se mantuvo así en la configuración de entrada de los archivos, para la generación de nueva información se dejó el mismo formato.

Para hacer el proceso de unificación de archivos se utilizaron varias herramientas como: SPSS, RapidMiner, SQLServer (Herramienta para importar y exportar) y la base de datos del mismo proveedor. En la siguiente ilustración se resumen el proceso que se realizó con los archivos y cómo se unirá con el resultado esperado por el Centro de Excelencia – CAOBA.



Teniendo en cuenta la ilustración anterior describiremos los diferentes repositorios.

1. **Repositorio inicial:** se propone este espacio para cargar directamente de la fuente de datos, de tal forma que la manipulación se facilite en los volúmenes de información que se pueden presentar, y servirá como repositorio de consulta de la información original que se cargó, permitiendo hacer validaciones de la integridad de la información y de las razones por las cuales dio como resultado en los pasos posteriores. En la ilustración se visualiza con el nombre "35 archivos planos"

2. Repositorio Unificado: se almacenará el resulta de transformar, limpiar y organizar la información del "Repositorio Inicial", unificando la información en una misma tabla, que permita comparar y validar la correlación y coherencia de la información de manera conjunta.
3. Vista minable: Resultado del proceso de verificación de calidad y limpieza para diseñar los modelos analíticos.
4. Repositorio de datos: en este repositorio se propone la organización de los datos para que se disminuya la redundancia de información y se optimice el almacenamiento, esto permitirá un mejor entendimiento de los datos y una mayor interoperabilidad y reutilización de estos. En este punto los metadatos asociados a cada variable deben estar claramente definidos y actualizados en el diccionario de datos, será la fuente de datos para los conjuntos de datos que se cargarán en la prueba prototipo.

7. Descripción de la Base de Datos

GENERAL

Tipo de Investigación:	Registro Administrativo.
Nombre de la Investigación:	Resultados del examen de estado realizado a los estudiantes en undécimo grado del colegio.
Entidad Responsable:	ICFES – Instituto Colombiano para la Evaluación de la Educación

Según el Decreto 869 de 2010, por el cual se reglamenta el examen de Estado de la Educación Media, ICFAESABER 11° [3]. Este examen tiene tres objetivos principales,

1. Seleccionar estudiantes para la educación superior
2. Monitorear la calidad de la formación que ofrecen los establecimientos de educación media
3. Producir información para la estimación de valor agregado de la educación superior.

En resumen, este examen inició en el año 1968 creado con el fin de seleccionar aspirantes a la Universidades, en 1980 el examen se vuelve obligatorio para el ingreso a la Educación Superior y se comienza la idea de hacer monitoreo a los establecimientos de la educación media, desde el año 2000 se reforma el examen para evaluar la calidad de la educación media y se crean 12 pruebas temáticas, en el año 2006 se integran historia y geografía a ciencias sociales y finalmente en el periodo 2014-2 se ajusta el examen para evaluar competencias comparables con el Saber PRO y de esta manera responder al tercer objetivo.

BASE DE DATOS

Periodo de Referencia:	2000 hasta el 2017
Total Registros:	9.285.527
Total Variables:	234
Cobertura:	Nacional, todos los departamentos

Se presenta como **Anexo 3**. Resumen de variables, donde se relacionan las 234 variables, nombre, descripción, tipo de variable (cadena, numérica, fecha, id) y se presenta la permanencia en el tiempo.

En la tabla 2, se muestra la distribución de los registros en cada uno de los semestres, la diferencia que se presenta entre los semestres 1 y 2, corresponde en gran medida a los estudiantes de undécimo de grado de colegios calendario A (semestre 1) y calendario B (semestre 2). En el año 2000 se presentaban aproximadamente 400.000 estudiantes, en el año 2016 se presentaron más 670.000.

Tabla 3. Distribución de registros por periodo

Periodo	Frecuencia	Porcentaje	Porcentaje acumulado
20001	77.626	0,8	0,8
20002	354.862	3,8	4,7
20011	77.961	0,8	5,5
20012	320.860	3,5	9,0
20021	63.371	0,7	9,6
20022	370.235	4,0	13,6
20031	69.736	0,8	14,4
20032	354.589	3,8	18,2
20041	64.779	0,7	18,9
20042	366.450	3,9	22,8
20051	72.842	0,8	23,6
20052	383.964	4,1	27,8
20061	61.346	0,7	28,4
20062	408.326	4,4	32,8
20071	69.303	0,7	33,6
20072	438.144	4,7	38,3
20081	68.868	0,7	39,0
20082	437.065	4,7	43,7
20091	69.743	0,8	44,5
20092	518.220	5,6	50,1
20101	30.321	0,3	50,4
20102	586.779	6,3	56,7
20111	86.674	0,9	57,6
20112	564.171	6,1	63,7
20121	95.946	1,0	64,7
20122	575.646	6,2	70,9
20131	86.470	0,9	71,9
20132	573.662	6,2	78,1
20141	88.685	1,0	79,0
20142	574.395	6,2	85,2
20151	108.258	1,2	86,4
20152	573.128	6,2	92,5
20161	74.127	0,8	93,3
20162	605.982	6,5	99,9
20171	12.993	0,1	100,0
Total	9.285.527	100,0	

Con relación a las variables que componen la base de datos, se identifican 234 variables que se agrupan en información personal, geográfico, socioeconómica del estudiante, información del colegio que registra la información, los resultados las competencias evaluadas y variables adicionales, tales como, Id de la base, id del estudiante para proceso de investigación.

Tabla 4. Descripción general de las variables de la base de datos

Grupo de variables	Variables					
	Total	Numéricas	Categorías	Binarias	Texto	Fecha
Información personal	24	2	8	13		1
Información de geoGráfico	20		18		2	
Información socioeconómica	118	3	64	44	2	5
Información del colegio	17	1	14		2	
Resultados	48	25	19		4	
Variables adicionales	7		5		2	
Total de variables	234	31	128	57	12	6

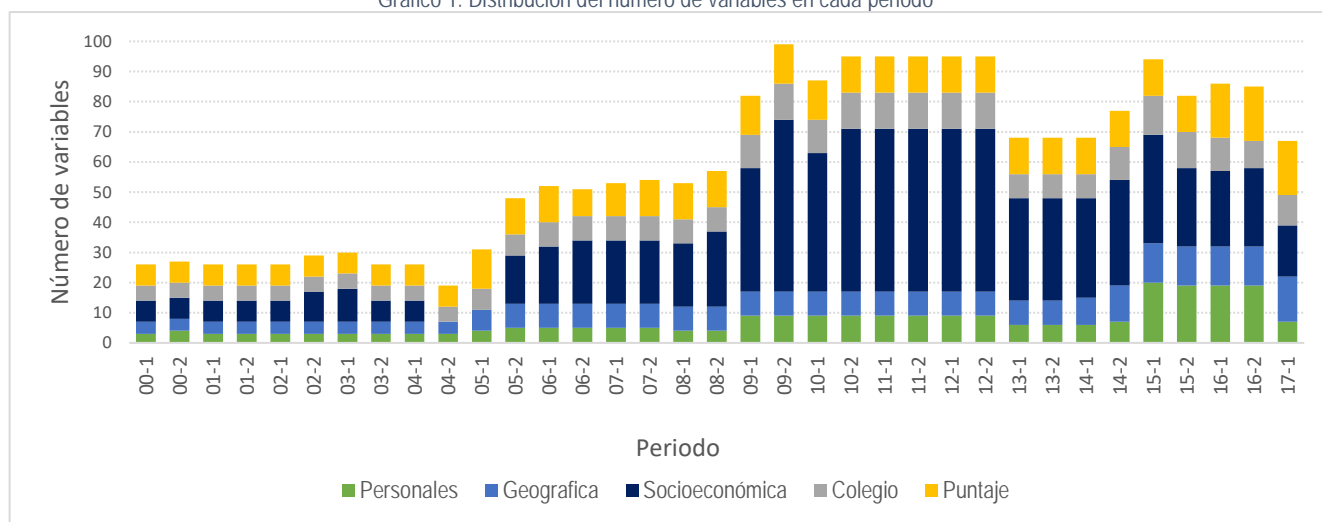
Fuente. Elaboración propia

En la siguiente Gráfico se presenta el número de variables registradas en cada semestre y su distribución en grupos temáticos de preguntas, tal como se presentó en la anterior tabla, donde se

identifica que desde 2000 hasta 2004 el número de variables no es mayor a 30, en el otro periodo 2005 hasta 2008 que las variables se duplicaron sobre todo en las variables socioeconómicas y finalmente desde 2009 hasta 2016 se registran alrededor de 90 variables.

Al compilar los 17 semestres de resultados, se evidenció que aparte del número de variables que se registraron también existían importantes cambios en las variables, las categorías, nombres y descripciones, por ejemplo, sólo en los años 2015 y 2016 se preguntaban por limitaciones física y mentales de los estudiantes en total 11 variables, y sólo en los años 2009 a 2012 se preguntaban por si el estudiante había reprobó cada curso en total 12 variables. Lo anterior, implicó que el número de variables totales de la base fuera superior a 200.

Gráfico 1. Distribución del número de variables en cada periodo



Fuente. Elaboración propia

8. Exploración de la Base de Datos

Las pruebas saber 11 han presentado durante estos 17 años grandes cambios en las competencias evaluadas, el cambio más importante se presenta en el periodo 2014-2 donde pasan de evaluar 8 a 5 componentes, la única prueba que permanece sin cambios en el tiempo es inglés.

Tabla 5. Competencias evaluadas en el ICFES desde 2000 hasta 2017

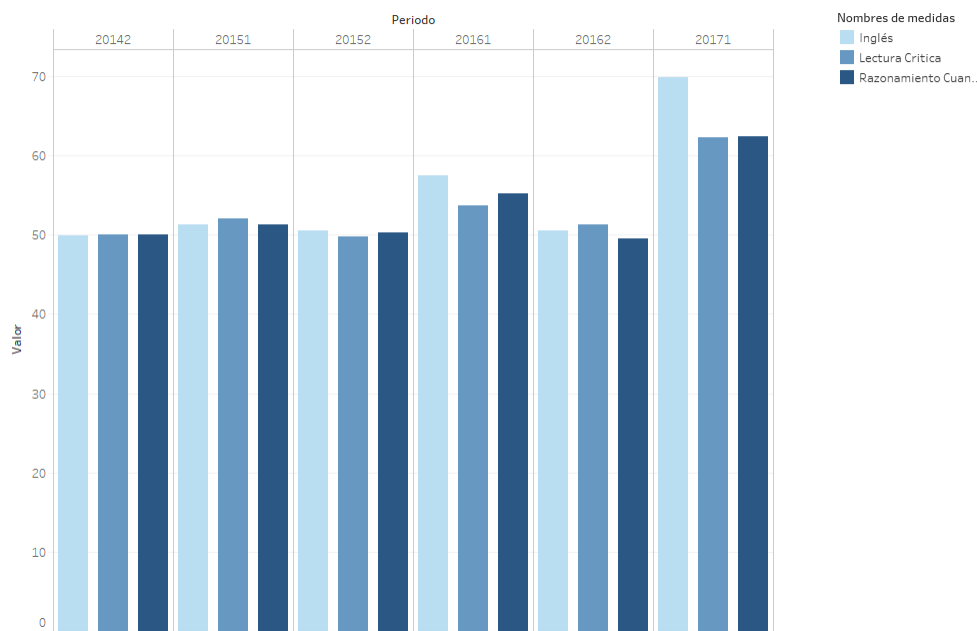
Hasta 2014-1	Desde 2014-2
Lenguaje	Lectura crítica
Matemáticas	Razonamiento cuantitativo
Biología	Ciencias naturales
Física	
Química	
Ciencias Sociales	Ciencias sociales y ciudadanas
Filosofía	
Inglés	Inglés

Fuente. Elaboración propia

En la Gráfico 2 se presenta la distribución del promedio del puntaje en inglés, razonamiento cuantitativo y lectura crítica en los últimos años, donde la escala de las pruebas no ha cambiado en el tiempo siendo de 0 a 100 con media 50 y desviación estándar 10; particularmente en el periodo 2017-1 se evidencia incremento del promedio del puntaje, esto se debe a que hasta el momento sólo han publicado los resultados de los estudiantes de colegio de calendario B, es decir, estudiantes que en general tienen mejores condiciones sociales y que la proporción con respecto al total de estudiantes que presentan en un año es del 5%, esto se debe tener presente para la toma de decisión de la base final para el modelo.

En relación a la *Meta 6*. estudiada, analfabetismo y competencias en aritmética de jóvenes se utilizarán las competencias, matemáticas y lectura de las pruebas saber 11, que según Murnane, Duhaldeborde y Tyler (2000) [4] sustentan que demostrar habilidades en lectura, pero especialmente en matemáticas, tiene un impacto significativo en el ingreso futuro de los estudiantes, además se consideran las competencias básicas en la gran parte de disciplinas.

Gráfico 2. Promedio de las competencias en los últimos años



Fuente. Elaboración propia

Con respecto a la variable edad, se evidencia que existen registros con datos inconsistentes, datos negativos o edades cercanas a 100, estos registros pueden afectar los resultados del análisis para lo cual se debe plantear una estrategia de imputación y limpieza. Adicionalmente en el gráfico, se muestra que en general no hay diferencia de edades entre géneros, sin embargo, si se evidencia el mismo efecto del calendario B mencionado anteriormente; en los primeros semestres el promedio de edad es menor que en el segundo. Específicamente se debe prestar atención a la calidad de la información en el año 2013, dado que su promedio es evidentemente superior que los demás años lo cual puede representar que ese año se registró la ausencia del dato con el número 999 (máximo de la tabla de descriptivos) y afecta directamente el promedio.

La variable edad está dada de manera explícita desde el año 2009 hasta el año 2017, el total de registros con esta variable son 5.223.475, con media igual a 19,27, la cual se considera que está perturbada por los atípicos. Los años anteriores registran la fecha de nacimiento lo cual permite el cálculo de esta variable.

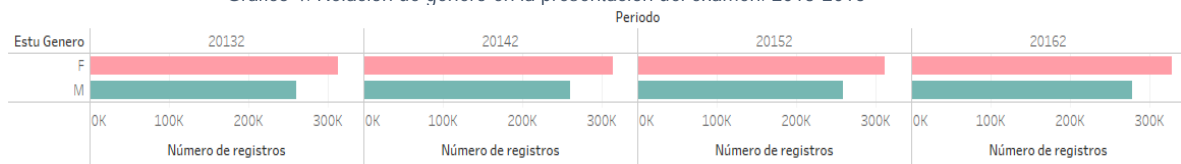
Gráfico 3. Distribución del promedio de edad por periodo y sexo



Fuente: Elaboración propia

Con relación a la variable género, la base total presenta datos faltantes en el 0,2% (22.922) de los registros mayoritariamente en los últimos años. El 54% de los registros corresponden a mujeres y el 45,7% a hombres, en el siguiente gráfico se muestra la relación de mujeres vs hombres en los segundos periodos de los cuatro últimos años, donde siempre las mujeres superan la presencia en las pruebas saber 11.

Gráfico 4. Relación de género en la presentación del examen. 2013-2016



Fuente: Elaboración propia

Particularmente, en el año 2016 se presentaron 680.109 personas, distribuidas en 366.066 mujeres y 313.760 hombres, en la Gráfico 5, se muestra la distribución por edad simple, donde se concentra el 61% de los registros en el rango de 16 a 17 años. A partir de las proyecciones de población⁴ realizadas por el Departamento Administrativo Nacional de Estadística (DANE) con base en el censo de población 2.005, es posible conocer el número de jóvenes en edades de 16 a 17 años en Colombia para el año 2016, esto con el fin de realizar un acercamiento a la proporción de participación de los jóvenes en las pruebas saber 11. En la tabla 5, se representa que el 24% de los jóvenes de 16 a 17 años en Colombia que deberían presentar la prueba finalmente la aplican, algunos son personas menores o mayores de este rango de edad, no obstante, la

⁴ <http://www.dane.gov.co/index.php/en/statistics-by-topic-1/population-and-demography/population-projections>

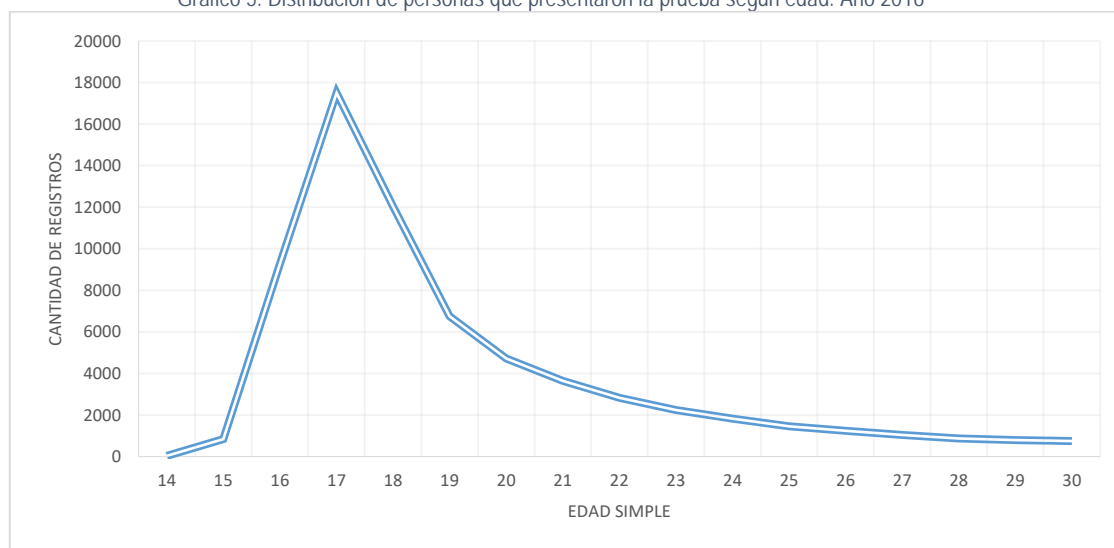
diferencia hace pensar que aún falta fortalecer la cobertura educativa, adicionalmente, teniendo el gráfico 4, existen mayor presencia de mujeres en la prueba, sin embargo, registran casi dos puntos porcentajes menos de proporción que los hombres.

Tabla 6. Proporción de jóvenes que presentan la prueba saber 11. Año 2016

	Proyecciones de población	61% de los registros totales 2016	Proporción de participación
Total nacional	1.726.770	414.866	24,0%
Hombres	880.986	223.300	27,9%
Mujeres	845.784	191.394	22,6%

Fuente. Proyecciones de población (denominador), registro saber 11 (numerador) y tasa calculada propia.

Gráfico 5. Distribución de personas que presentaron la prueba según edad. Año 2016



Fuente. Elaboración propia

Dentro de las hipótesis que se plantean en este análisis es que las variables socioeconómicas y del colegio afectan los resultados académicos de cada grupo, en primera medida fue importante observar la distribución de la base de datos con respecto al estrato, el cual corresponde al estrato que registra la vivienda donde reside en el momento de la aplicación del examen. Aunque la pregunta *estrato de la vivienda* se ha presentado estable desde el año 2005 hasta el 2016, si registra el 28,3% de registros faltantes, no obstante, más del 90% de estos faltantes corresponde al periodo 2013-2. Ver Gráfico 7.

A partir de los registros de Colombia desde 2006-1 hasta el 2016-2 se logra identificar que el 38% de las personas que participan en el examen saber 11, pertenecen al estrato 1, seguido del 36,5% de estrato 2, es decir, que más 70% de la población colombiana que asistió a la educación secundaria en este periodo, son de los estratos socioeconómicos más bajos, al realizar la comparación con "*Estratificación socioeconómica para servicios públicos domiciliarios*" realizado por el DANE⁵, en el año 2016, el 66,1% de las viviendas era estrato 1 y 2. En la siguiente Gráfico se puede observar la distribución mencionada anteriormente, así como la proporción por género

⁵ <http://www.dane.gov.co/index.php/servicios-al-ciudadano/servicios-de-informacion/estratificacion-socioeconomica#generalidades>

en cada estrato, donde las mujeres representan en todos los estratos la mayor parte, aunque la diferencia es pequeña.

Gráfico 6. Distribución por género y estrato de las personas que presentaron la prueba. 2006-2016

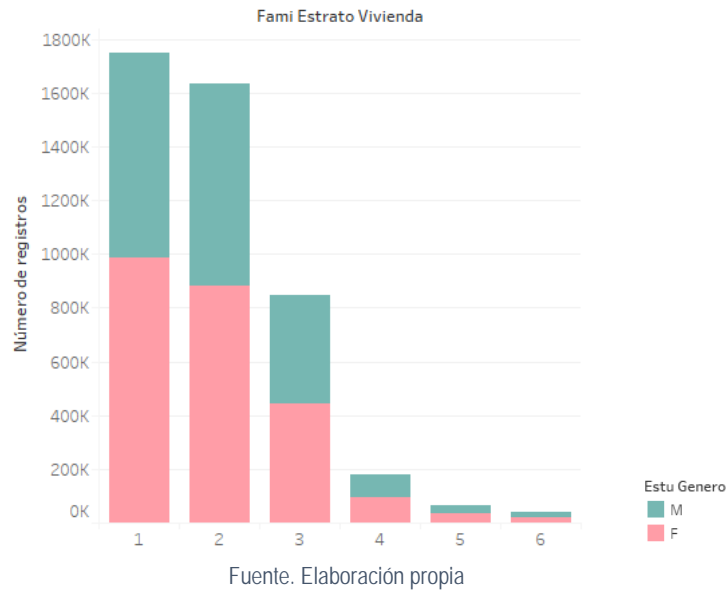
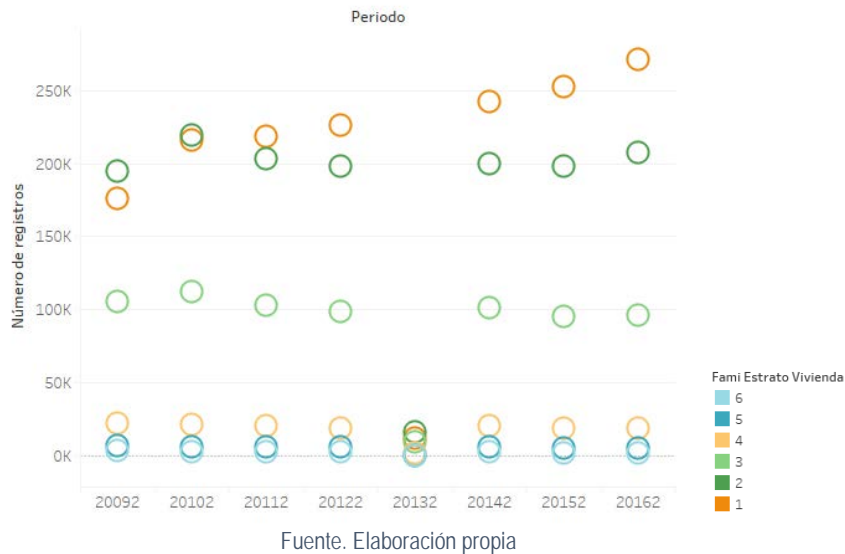


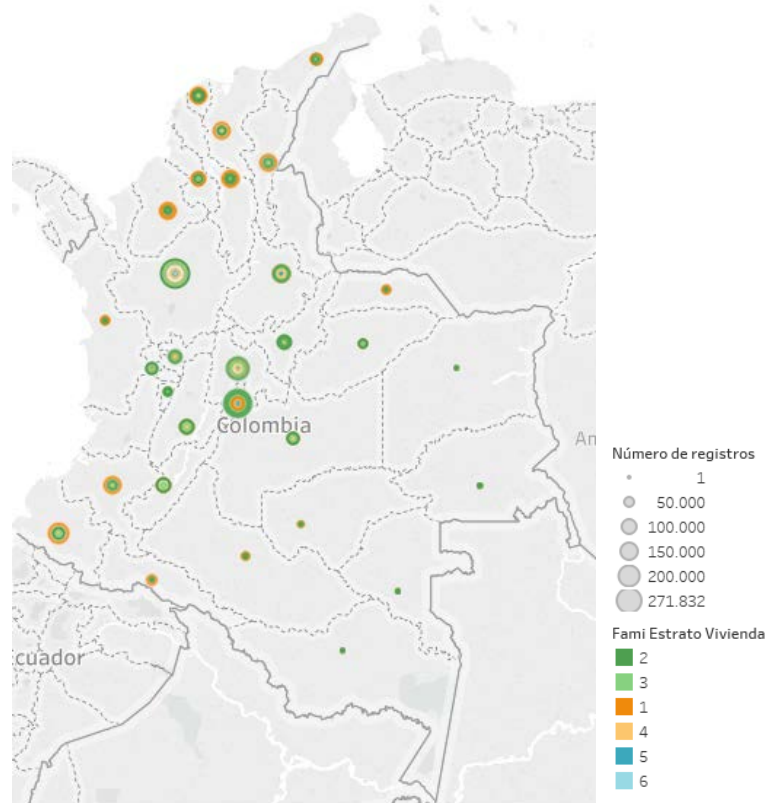
Gráfico 7. Distribución por género y estrato de las personas que presentaron la prueba. 2009-2016



Dentro de las variables con mayor calidad y utilidad de la base de datos son las geográficas, el ICFES indaga y/o registra el departamento y municipio donde reside el estudiante, así como donde presentó el examen y donde está ubicado el colegio, un posible análisis que se puede realizar es la movilidad de los estudiantes con motivo de presentación del examen. Teniendo en cuenta el departamento donde la persona presentó la prueba, es posible identificar que existe información para todos los departamentos del país. En el mapa se muestra que según el tamaño del círculo que Bogotá, Antioquia y Cundinamarca son las regiones donde más se presentan las personas,

adicionalmente, según los colores de los círculos se presenta la distribución por estrato en cada uno, se resalta que los departamentos que se rodean de verde significa que tiene mayor porcentaje de estrato 2, a diferencia de los departamentos rodeados de naranja significa que su mayor presencia es de personas con estrato 1.

Gráfico 8. Mapa Colombia, personas que participaron en la prueba por estrato. 2009-2016



Fuente. Elaboración propia

Otro aspecto de relevancia para el análisis es el nivel educativo y ocupación de los padres, específicamente estas cuatro variables registran información desde el año 2000, sólo en tres periodos no registraron información, lo cual representa aproximadamente el 17% de datos faltantes en cada variable para el total de registros.

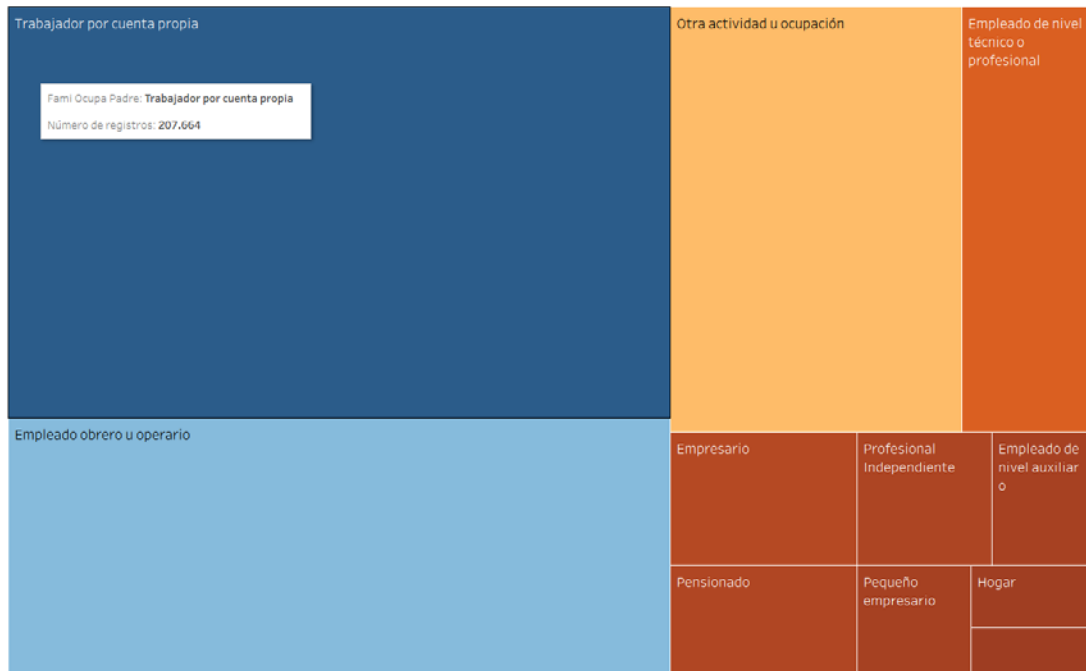
En la Gráfico 9 se decidió presentar la distribución del nivel educativo de la madre para las personas que presentaron el examen en el año 2016-2, donde se encontró que la tercera parte de las madres han terminado el bachillerato y no continuaron con sus estudios, seguido del 16% de las madres que no terminaron el bachillerato, por otro lado, en la Gráfico 10 se presenta distribución de la ocupación de los padres, el 38% se dedican a trabajar por su cuenta, seguido del 24% que trabajan por cuenta propia.

Gráfico 9. Distribución del nivel educativa de la madre en el año 2016-2



Fuente. Elaboración propia

Gráfico 10. Distribución de la ocupación del padre en el año 2016-2



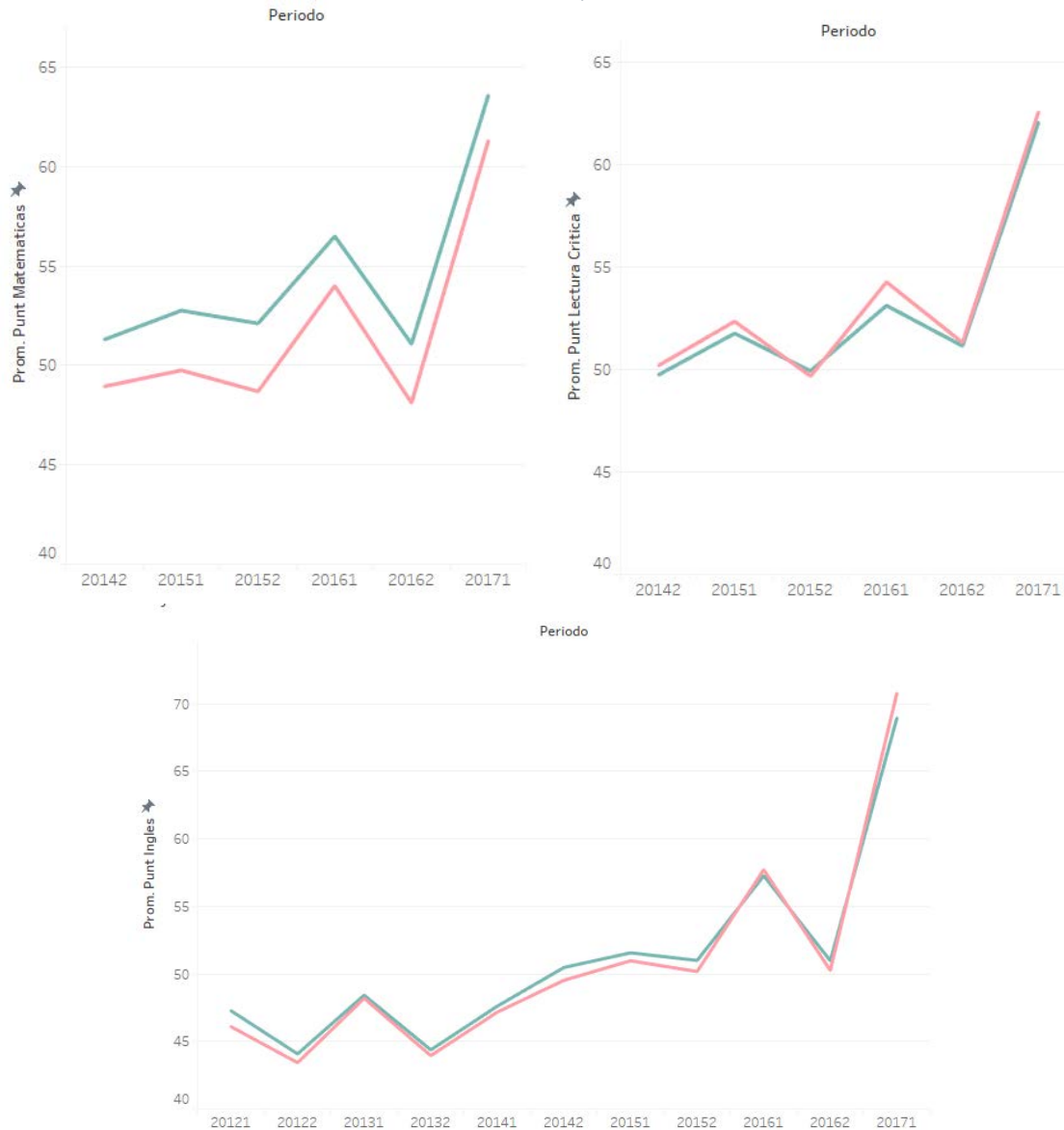
Fuente. Elaboración propia

Adicionalmente, el análisis de los puntajes de las competencias, lectura crítica, matemáticas e inglés, en la Gráfico 11, se presenta la distribución de los puntajes en estas competencias en los tres años por género, se evidencia una brecha entre géneros en los puntajes de las pruebas saber

11. Un primer resultado de este gráfico es que existe un comportamiento particular en el último periodo 2017-1, cabe resaltar que corresponden a estudiantes de calendario B, no obstante, presenta una diferencia importante con respecto al anterior grupo de calendario B, en este sentido, una de decisiones de la exploración es no utilizar este periodo para la aplicación del modelo.

Por otro lado, se observa que en la competencia de matemáticas los hombres tienen mejores resultados que las mujeres; visualmente esta brecha se está reduciendo. En la prueba de lectura crítica se observa que las mujeres tienen mejores resultados, pero la diferencia es mínima y con relación a inglés, mantienen un comportamiento similar los dos géneros, no obstante, los promedios están más bajo en inglés que en las demás competencias.

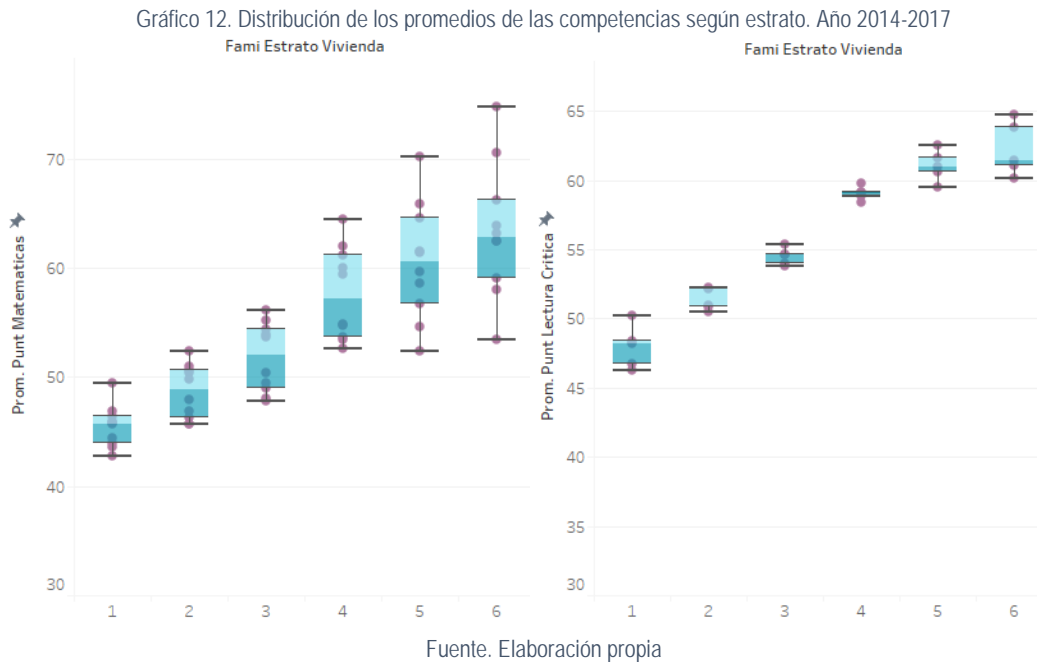
Gráfico 11. Distribución de los puntajes por sexo. Año 2014. 2017



Fuente. Elaboración propia

Otro resultado que llamó la atención de la exploración fue la presentada en la Gráfico 12, se muestra en un gráfico de caja, la relación del puntaje promedio de matemáticas y lectura crítica vs el estrato de la vivienda de la persona que presentó el examen, el resultado muestra una importante diferencia de los resultados según el estrato socioeconómico, en matemáticas hay mayor dispersión de los puntajes por estrato, en relación a lectura crítica donde la caja y los bigotes son más concentrados.

Relacionando estos resultados con la distribución por estrato presentada anteriormente, hace pensar que Colombia necesita gran trabajo en temas educativos en el nivel secundaria, puesto que la mayor parte de Colombia registra lo puntajes más bajos.

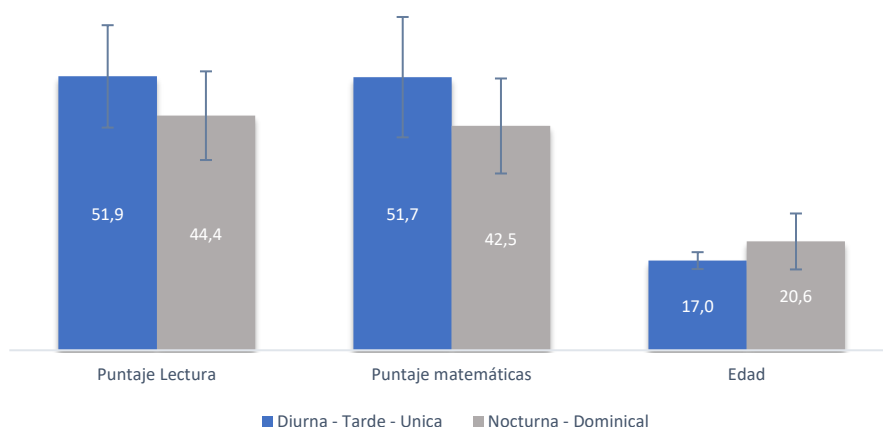


Se realizó la revisión de investigaciones desarrolladas en el marco de la equidad de género y el efecto en los resultados académicos, específicamente con base en la misma fuente de datos utilizada en este proyecto; se identificó el proyecto "Brechas de género en el rendimiento escolar a lo largo de la distribución de puntajes: evidencia pruebas saber 11°" realizado por dos profesoras de la Pontificia Universidad Javeriana, publicado en la revista *Vniversitas Económica* del departamento de Economía de la Facultad Ciencias Económicas y Administrativas, este estudio encontró una brecha significativa en matemáticas y ciencias en contra de las niñas y una brecha pequeña y significativa a favor de las niñas en lectura [5], información que se pudo observar en la anterior exploración de la base de datos.

No obstante, de una las principales conclusiones de este estudio que utilizó la técnica de descomposición de Juhn, Murphy y Pierce, es que la brecha identificada por sexo no es explicada por las características observadas, personales, familiares y escolares observables, el mayor de porcentaje de la brecha se atribuye a factores no observables en la base de datos, tales como, motivación, expectativas, ambiente escolar y culturales determinantes del rol de la mujer) [5].

En este sentido, se exploraron otras variables que permitieran identificar diferencias entre los puntajes de las pruebas. En el gráfico 13 se identifica que la jornada del colegio presenta resultados diferenciados en los puntajes, puesto que los estudiantes que provienen de colegios de jornada nocturna o dominicales registran menores puntajes, así mismo, son estudiantes con otras características socioeconómicas, por ejemplo, la edad de estos estudiantes es promedio es mayor a 20 años en cambio los estudiantes de jornada diurna, tarde o única tienen en promedio 17 años con una varianza más pequeña.

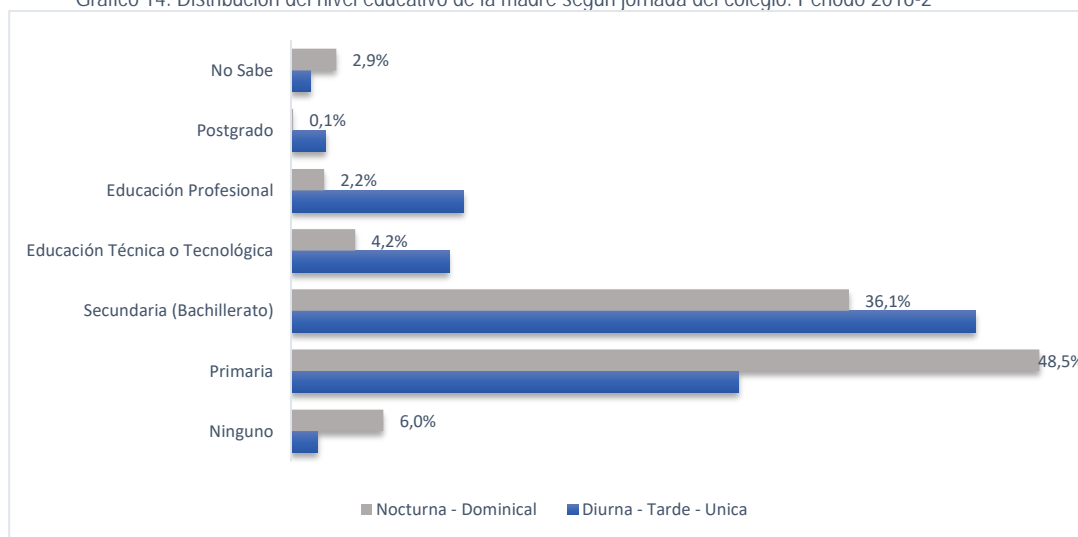
Gráfico 13. Distribución de los promedios de las competencias y edad según jornada del colegio. Periodo 2016-2



Fuente. Elaboración propia

En este sentido, otra variable interesante que logra ver la influencia de la jornada del colegio, es el nivel educativo de la madre, en el gráfico 14 se muestra que cerca del 50% de los estudiantes que provienen de colegio nocturno, la madre alcanza nivel educativo primaria, así mismo, también tiene mayor porcentaje de sin nivel educativo.

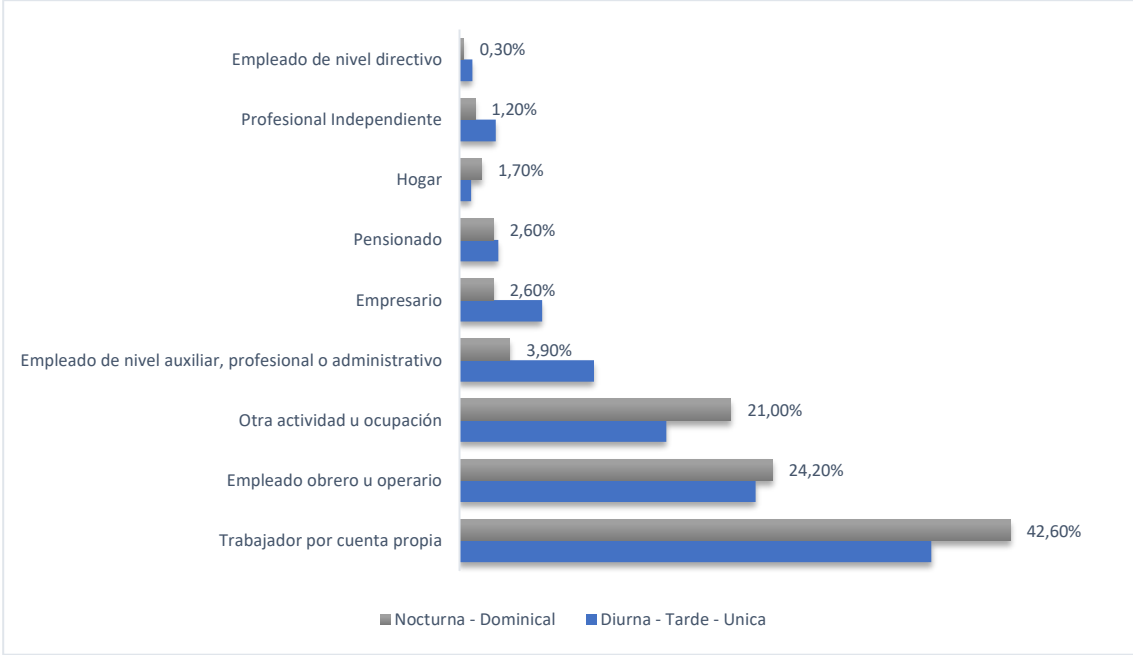
Gráfico 14. Distribución del nivel educativo de la madre según jornada del colegio. Periodo 2016-2



Fuente. Elaboración propia

El gráfico 14, muestra la ocupación de los padres de los estudiantes según la jornada de estudio, más del 40% trabajan por cuenta propia es decir es independiente pero no es profesional, seguido del 24% que son obreros y operarios, sin embargo, los padres de los estudiantes que provienen de colegios diurnos también registran las dos anteriores ocupaciones como las más frecuentes, en total suman más del 50% en estas actividades, esto va muy relacionado con el estrato socioeconómico que se ha mencionado anteriormente, puesto que más del 50% de los estudiantes que presentan el examen son de estrato 1 y 2.

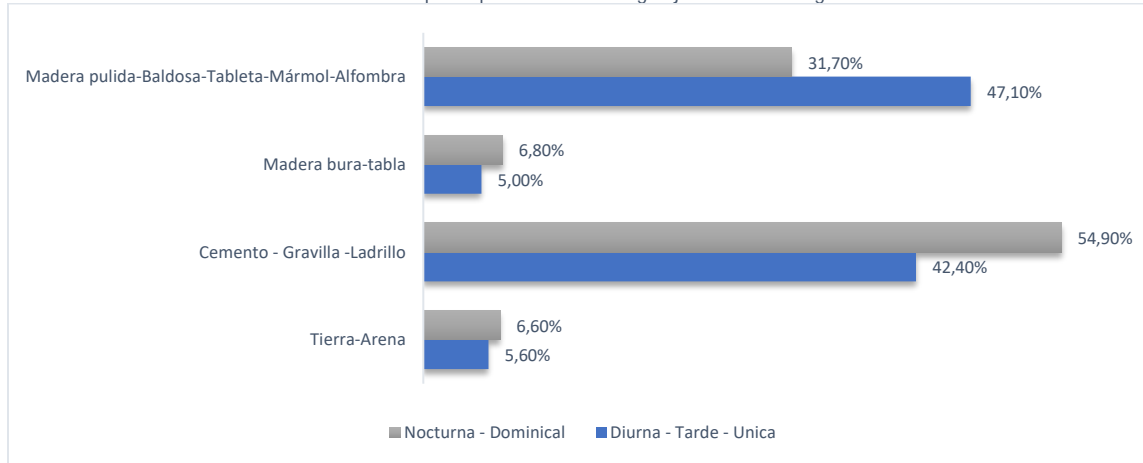
Gráfico 15. Distribución de la ocupación del padre según jornada del colegio. Periodo 2016-2



Fuente. Elaboración propia

Dentro de las variables que se registran en la base de datos, se encuentran algunas que pueden tratarse como variables proxy para definir que el estudiante vive en condiciones de pobreza. La metodología de NBI busca determinar, con ayuda de algunos indicadores simples, si las necesidades básicas de la población se encuentran cubiertas. Los grupos que no alcancen un umbral mínimo fijado, son clasificados como pobres. Los indicadores simples seleccionados, son: Viviendas inadecuadas, hogares con hacinamiento crítico, viviendas con servicios inadecuados, hogares con alta dependencia económica y hogares con niños en edad escolar que no asisten a la escuela.[6] En el censo de 2005, se registró que el 10,5% de la población viven en viviendas inapropiadas y una de las variables que definen esa vivienda inapropiada es la calidad del piso; tal como se presenta en la siguiente Gráfico, el 6% de los estudiantes que estudian en jornada nocturna viven en casa con piso de tierra y arena.

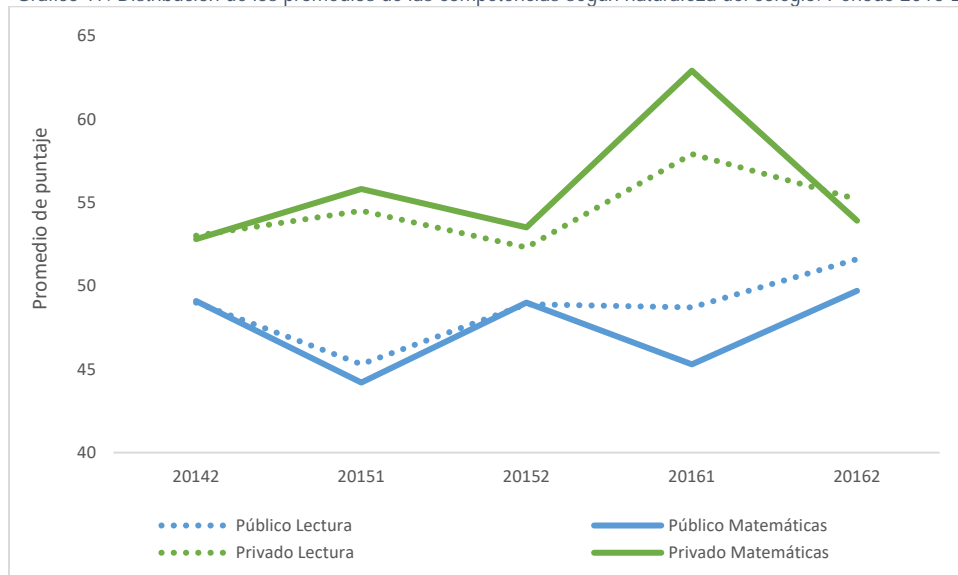
Gráfico 16. Distribución del tipo de piso de la casa según jornada del colegio. Periodo 2016-2



Fuente. Elaboración propia

Otra variable que genera diferencias importantes en los resultados de saber 11, es la naturaleza del colegio, es decir, si es público o privado, en el periodo 2016-2 la distribución de estudiantes por naturaleza de colegio fue 22,7% privado y 77,3% públicos, adicionalmente, si se observa la Gráfico 18, se puede observar la marcada diferencia en los puntajes de las pruebas lectura crítica (puntos) y matemáticas (línea), por la naturaleza, los colegios públicos (azul) hasta el periodo 2016-1 presentan en promedio menos de 50 de puntaje y se observa crecimiento en el último periodo de análisis, por otro lado, los colegios privados (verde) están sobre 50, pero el periodo 2016-1 muestra un incremento que se debe analizar con mayor detalle.

Gráfico 17. Distribución de los promedios de las competencias según naturaleza del colegio. Periodo 2016-2



Fuente. Elaboración propia

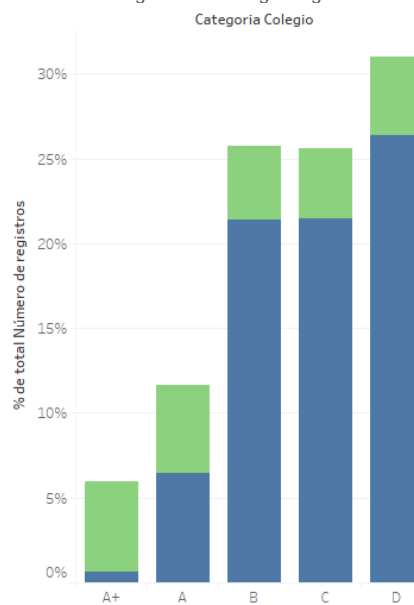
A partir del periodo 2014-2 el ICFES cambió la metodología para clasificar los planteles educativos a partir de los resultados de las pruebas saber 11⁶; estas clasificaciones fueron posibles descargarlas para los últimos periodos y realizar el respectivo cruce con la base de datos manejada en este proyecto. El ICFES clasifica a los colegios en 5 categorías descritas de la siguiente manera:

Tabla 7. Clasificación colegios. ICFES 2014-2

Categoría	Descripción
A+	Tienen más del 85% de sus estudiantes en el 33% superior, en al menos una de las cinco pruebas
A	Los establecimientos A tienen más del 65% de sus estudiantes en el 33% superior, en al menos una de las cinco pruebas.
B	Los establecimientos B tienen entre el 40% y el 85% de sus estudiantes en el 33% superior, en al menos una de las cinco pruebas.
C	Los establecimientos C tienen entre el 15% y el 60% de sus estudiantes en el 33% superior, en al menos una de las cinco pruebas.
D	Los establecimientos D tienen menos del 40% de sus estudiantes en el 33% superior, en al menos una de las cinco pruebas.

En relación con lo anterior, los colegios A+ se consideran los mejores clasificados del país, sin embargo, al observar la Gráfico 18 se evidencia que el porcentaje de colegios A+ no supera el 6%, adicionalmente, en color verde se presenta la proporción de colegio privado y azul los colegios públicos, esto es otra evidencia que existe una marcada diferencia entre la naturaleza de los colegios. El 30% de los colegios están clasificados en la categoría más baja, donde más la gran parte son colegios públicos. En el prototipo del observatorio se presentarán los agregados de colegio por categorías por departamento.

Gráfico 18. Distribución de las categorías de colegio según su naturaleza. Periodo 2015-2



Fuente. Elaboración propia

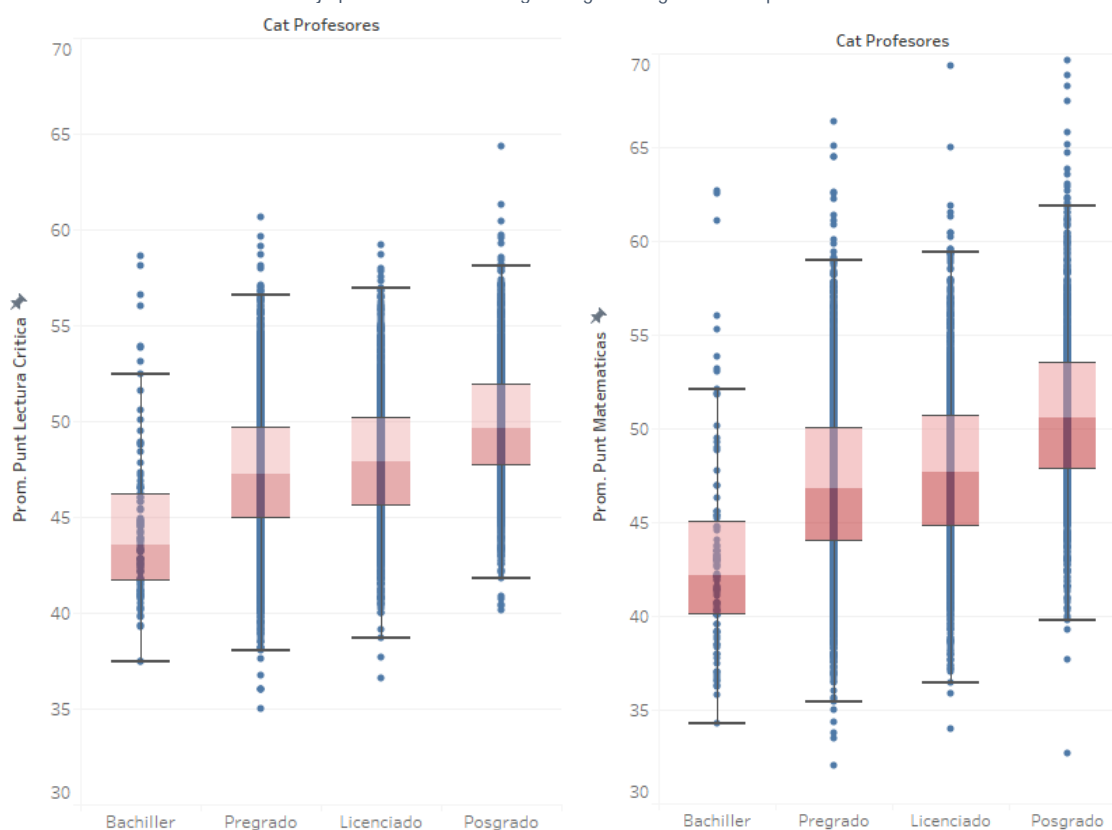
⁶ <http://www2.icfes.gov.co/instituciones-educativas-y-secretarias/saber-11/clasificacion-de-planteles>

Para terminar la exploración de variables interesantes del contexto educativo, en la página de datos abiertos⁷ se logró obtener información de la planta docente de los colegios para el periodo 2015-2, esta base relaciona la cantidad de docentes por sexo y nivel educativo. Se logró cruzar con más 75% de la base de resultados de saber 11 del periodo 2015-2. Se generó una variable que se pueda interpretar de la siguiente manera:

- Bachiller, hace referencia a los colegios donde más del 50% de sus docentes tienen este como título máximo de educación.
- Pregrado, más del 50% de sus docentes son profesionales no licenciados
- Licenciado, más del 50% de sus docentes son profesionales licenciados
- Posgrado, más del 50% de sus docentes tiene posgrado en el área de educación u otra.

En el gráfico 19, se presenta un resultado interesante, puesto que se evidencia una relación entre los puntajes promedio de los colegios con respecto al tipo de profesores que lo caracterizan. Al lado izquierdo se presenta el promedio del puntaje en lectura crítica, donde los colegios con mayoría de profesores bachilleres quedan en los promedios más bajos y los colegios con profesores con posgrado tienen mayor promedio. La limitación es que esta variable sólo fue posible cruzar para un periodo de análisis, no obstante, se comparará sus resultados e importancia en los modelos.

Gráfico 19. Puntaje promedio de los colegios según categoría de los profesores. Periodo 2015-2



⁷ <https://www.datos.gov.co/>

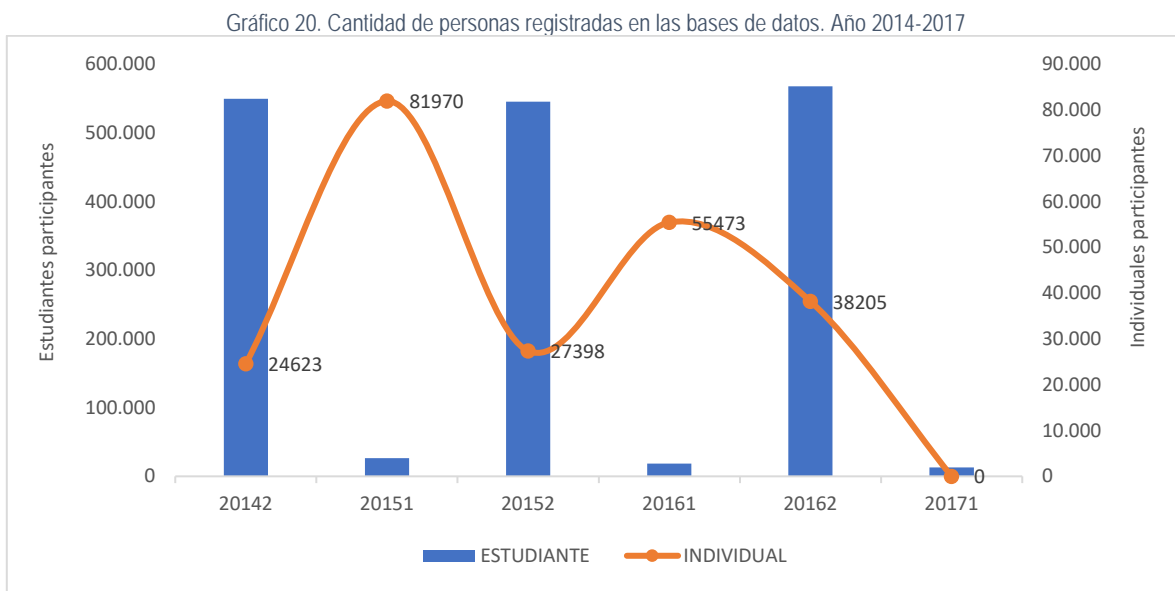
9. Verificación de la Calidad de los Datos

A partir de la exploración realizada a la base de datos, se identificaron los siguientes aspectos de problemas de calidad y/o limitaciones de la información:

- En la Gráfico 1, se presenta que la cantidad de variables en el tiempo no es constante, esto implica que se deben limitar las variables del análisis. No obstante, desde el año 2014-2 se presenta mayor consistencia de las variables.
- En relación a las variables geográficas, gran parte de éstas presentan alto nivel de missing lo cual dificulta el análisis. La variable con mejor resultado con respecto a la completitud de la información es, ESTU_DEPTO_PRESENTACION, es decir el departamento donde el estudiante presentó el examen, sin embargo, se evidencia que por ser una variable en texto tiene problemas con la Ñ y tildes lo cual debe corregirse.
- Para la variable género se presentan en los años 2002 y 2003 dos categorías diferentes a femenino y masculino las cuales deben corregirse. Sin embargo, durante los 17 años de estudio esta variable presenta buena calidad de información, el máximo periodo con datos faltantes es el 2014-1 donde el 1,4% de los registros no indican el género, en este caso, se eliminaron estos registros.
- La exploración de datos también permitió identificar que el 20% de las variables (47) no registran información en ningún periodo, están totalmente en blanco.
- Se presentan variables como, fecha de nacimiento, mes de nacimiento y año de nacimiento, las cuales requieren de unificación. En este mismo sentido, variables que en el tiempo presenta cambios pequeños de nombres.
- Las características del hogar referente a los bienes que posee, en unos años se registraron con S y N y otros años como 1 y 0, se realizó la respectiva unificación.
- La variable que indica si el estudiante trabajaba durante sus estudios ha cambiado en el tiempo, hasta de 2014 la variable registraba 7 opciones de respuestas y ahora son tres: no trabaja, si más de 20 horas, si menos de 20 horas. Se agrupó para volver dicotómica y comparable: trabaja o no trabaja. La variable etnia se agrupó de la misma manera, pertenece o no a una etnia.
- Como se presentó en la Gráfico 7, en el año 2013 la variable estrato socioeconómico tiene alto porcentaje de valores faltantes. Se eliminaron estos registros.
- Las variables de puntaje antes del 2009 presentan alta inconsistencia, valores perdidos y rangos incoherentes. Tal como se mencionaba en la fase anterior, este examen ha presentado importantes cambios en el transcurso del tiempo, debido a sus objetivos y el uso de

comparabilidad con la prueba SaberPRO que le permita generar el análisis de valor agregado de las instituciones de educación superior.⁸ Se decidió utilizar la información desde el año 2014-2 con el fin de analizar los puntajes de las pruebas que permitan conclusiones acertadas y que no se afecten por cambios propios del examen.

- Por otro lado, se evidencia que la base 2017 está incompleta dado que no registra resultados para las personas que se presentan de manera individual, tal como se observa en la línea naranja, en los primeros periodos corresponde a más del 50% de los registros. El eje izquierdo muestra el rango de personas que se presentan como estudiantes el cual es superior a 500.000, el eje derecho presenta las personas que se presentan de forma individual el cual no supera los 90.000, sin embargo, en el año 2017 registra en cero y esto evidencia problemas de completitud de la base.

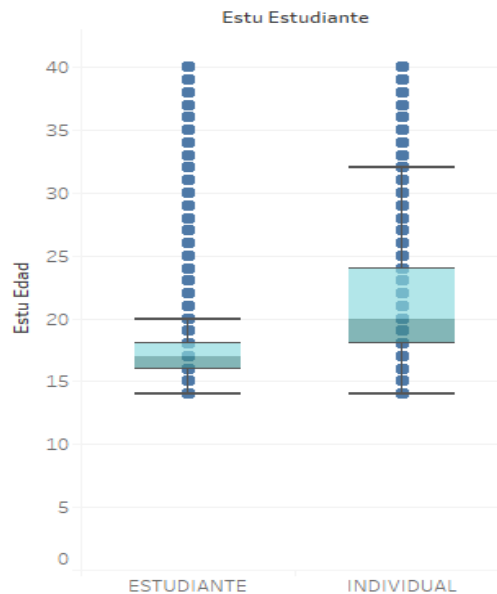


- El año 2017 presenta un cambio importante en el número de variables que registra, 10 variables se eliminaron y se crearon otras 7, acerca de la alimentación del estudiante, dedicación a internet, nacionalidad del estudiante y nuevos bienes de la casa. Esto también dificulta la comparabilidad.
- En el año 2017 las variables de ocupación y nivel educativo del padre y la madre cambiaron con respecto a los anteriores años.
- Las características de las personas que presentan el examen como estudiantes o como individuales tienen algunas diferencias importantes, por ejemplo, en el gráfico se puede observar la distribución de la edad de los grupos, evidentemente la edad de las personas que se presentan de manera individual, tiene mayor promedio de edad y presentan mayor rango de edades, a diferencia

de los estudiantes que tiene promedio entre 16-20 años, así mismo se observan datos atípicos que se deben ajustar. Ver Gráfico 20.

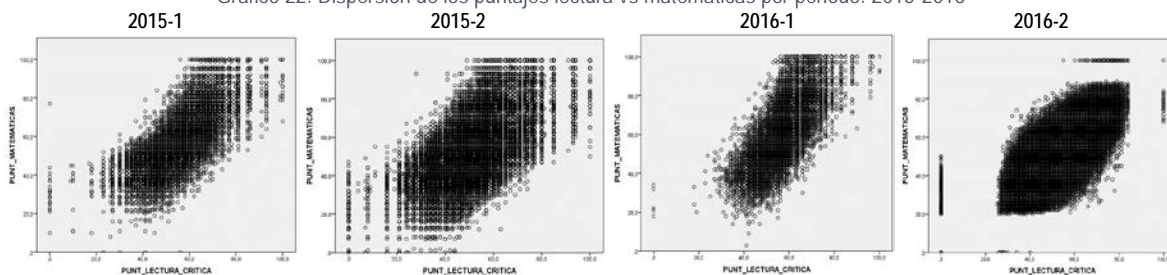
- Dado que se analizarán también el efecto de las características del colegio en los puntajes de las pruebas, se decidió filtrar la base para las personas que se presentan como estudiantes de colegio, puesto que las otras no tienen información de colegio y nos genera muchos vacíos propios de la base. Se propone en otros estudios trabajar comparar los resultados de los grupos y de esta manera identificar las brechas entre ellos.
- Para las técnicas analíticas es importante revisar la dispersión de los puntajes de lectura vs matemáticas, en el gráfico 21 se muestra la dispersión de cuatro periodos, donde se identifican algunos atípicos que pueden afectar la calidad de las técnicas por ende se ajustaron. Adicionalmente, se evidencia un aumento en la concentración de los resultados en el último periodo de análisis.

Gráfico 21. Distribución de edad por tipo de presentación. Año 2014-2016



Fuente. Elaboración propia

Gráfico 22. Dispersión de los puntajes lectura vs matemáticas por periodo. 2015-2016



Fuente. Elaboración propia

FASE 3. PREPARACIÓN DE LOS DATOS

El objetivo de esta fase es construir el conjunto de datos final a partir de datos brutos que serán la entrada para la herramienta de modelado.

10. Seleccionar la Base de Datos

Para la selección de las variables se tuvieron en cuenta los siguientes criterios:

- Periodo de tiempo con el número mayor de variables compartidas,
- Variables con más del 80% de la información completa en el periodo seleccionado,
- Variables con categorías coherentes de acuerdo con la descripción dada en el diccionario de datos.
- Variables con mayor relevancia para resolver las hipótesis planteadas, a partir de lo presentado en la exploración de la base de datos.

Tal como se mencionó anteriormente, desde el año 2014-2 se evidencia mayor consistencia de las variables en el tiempo, no obstante, en el año 2017 se presenta un cambio importante de variables nuevas y eliminación de algunas que venían coherentes en el tiempo, para lo cual se decidió utilizar periodo 2014-2-2016-2. En la tabla 5 se presenta la distribución por categorías de las variables eliminadas y las nuevas variables.

Tabla 8. Relación de número de variables eliminadas

Grupos de variables	# Inicial	# Eliminadas	Nuevas
Información personal	24	19	1
Información de geoGráfico	20	14	
Información socioeconómica	118	97	
Información del colegio	17	11	
Resultados	48	41	2
Variables adicionales	7	6	

Fuente. Elaboración propia

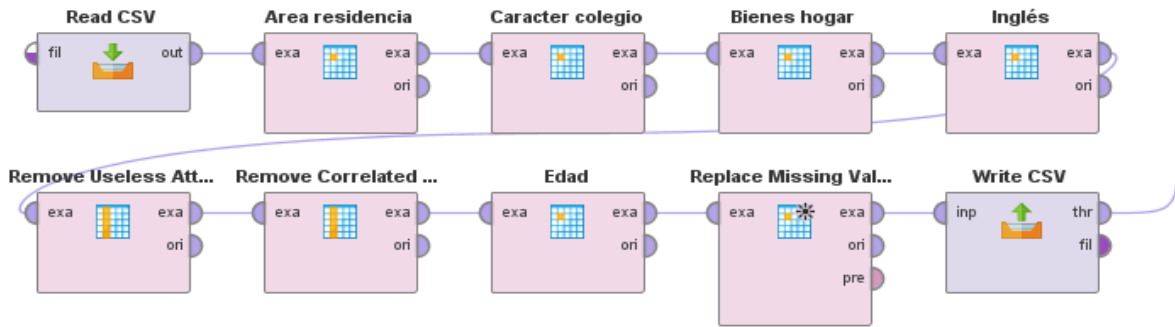
BASE DE DATOS

Periodo de Referencia:	2014-2 hasta el 2016-2
Registros:	1.685.989
Variables base ICFES:	47
Cobertura:	Nacional, todos los departamentos

11. Limpieza de la Base de Datos

El resultado del análisis exploratorio permite identificar variables con baja relevancia por la completitud de los datos o la cantidad de información que aportaban al análisis, así como datos faltantes, atípicos e incoherentes. Con la ayuda del software RapidMiner se realizó la limpieza a la base de datos, sistematizada en el siguiente flujo. A continuación, se relacionan los criterios y decisiones.

Ilustración 4. Flujo de trabajo en RapidMiner



11.1. Variables que no aportaban información:

Las variables en las cuales una sola categoría acumulaba el 95% del total de registros, o cuya categoría con mayor frecuencia acumulaba 5% o menos y/o las variables constantes, se eliminaron de la base de datos. Para ello se utilizó el procedimiento *Remove Useless Attributes*. Como resultado se eliminaron las variables fecha de nacimiento, nombre de colegio.

11.2. Variables redundantes

Se eliminaron variables correlacionadas de la base de datos a partir del procedimiento *Remove Correlated attributes*, los parámetros definidos fueron, 95% de correlación o más, como resultado se reconocieron las siguientes variables correlacionadas: puesto del estudiante y percentil de inglés.

Adicionalmente, de acuerdo con el análisis exploratorio presentado anteriormente se identifica completa relación entre las variables departamento y municipio, en este sentido, se decide eliminar el código del municipio dado que no aporta información adicional y se pretende prever sobreajuste del modelo.

11.3. Registros inconsistentes e incompletos (imputación)

En la variable edad, el 5% de los registros presentaban valores inconsistentes (edades menores a 14 años y mayores a 80 años). Teniendo en cuenta que estos registros contaban con información para la mayoría de las variables restantes, y dado que la edad de los estudiantes presenta una distribución muy concentrada entre los 16 y 18 años, se procede a imputar los valores con la mediana del atributo.

Las variables de puntaje en matemáticas y lectura fue necesario realizar una revisión de datos atípicos, con ayuda del programa SPSS y el comando *DETECTANOMALY*, el cual busca casos atípicos basados en desviaciones de las normas de sus agrupaciones, el procedimiento está diseñado para detectar rápidamente casos atípicos con fines de auditoría de datos en el paso del análisis exploratorio de datos, antes de llevar a cabo cualquier análisis de datos inferencial. En la tabla 7, se presenta los registros atípicos identificados y eliminados de la base.

Tabla 9. Número de registros atípicos por semestre y por competencia. Año 2014-2 – 2016-2

	2014-2	2015-1	2015-2	2016-1	2016-2
Total antes de limpieza	522.821	22.924	514.525	15.700	521.572
Puntaje lectura	12.281	584	12.172	558	12.592
Puntaje matemáticas	13.860	562	13.554	227	13.487
Total después de limpieza	496.680	21.778	488.799	14.915	495.493

Fuente. Elaboración propia

FASE 4. MODELAJE

En esta fase de la metodología se escogerá la técnica (o técnicas) más apropiadas para los objetivos definidos de analítica. A continuación, y una vez realizado un plan de prueba para los modelos escogidos, se procederá a aplicar dichas técnicas sobre los datos para generar el modelo y por último se tendrá que evaluar si dicho modelo ha cumplido los criterios de éxito o no.

12. Técnicas de modelado seleccionadas

Para la selección de las técnicas que se aplicarán es necesario identificar las más apropiadas para el tipo de problema que se desea resolver y se debe considerar el objetivo del proyecto en términos analíticos. En este sentido, como se ha mencionado en la primera fase del entendimiento del negocio, el Centro de Excelencia CAOBA plantea en uno de los objetivos construir y desarrollar un modelo analítico con el propósito de convertir las fuentes disponibles en los datos abiertos en temas de Educación en Colombia en información útil con miras a proporcionar análisis de valor en el país y de esta manera apoyar los indicadores de los Objetivos de desarrollo sostenible.

La fuente de datos seleccionada para el desarrollo de los objetivos analíticos se ha encontrado que responde a algunos indicadores mencionado en el **Producto No.1 Caracterización de indicadores de los ODS**, en este sentido se considera que contiene información relevante sobre la educación media en Colombia. Actualmente, dentro de la búsqueda del incremento en la cobertura en educación media en todos los niveles se pueden descuidar aspectos de interés tales como, ¿Es posible generar conglomerados de estudiantes a partir de las competencias de lectura y matemáticas?, ¿cuáles son las variables que más afectan el desempeño de los estudiantes en las pruebas saber 11?

De acuerdo con lo anterior, se plantea desarrollar dos técnicas analíticas:

12.1 Clustering

Se busca agrupar a los estudiantes en tres niveles de desempeño (bajo, medio y alto) de acuerdo con sus puntajes en matemáticas y lectura, para esto aplica la técnica de clustering, la cual permite evidenciar los grupos naturales que surgen de los resultados de las pruebas saber 11. Este resultado permitirá la aplicación de la segunda técnica analítica.

Aspectos a tener cuenta:

- Las variables incluidas en este modelo son los puntajes de lectura y matemáticas, es decir numéricas.
- Las variables de puntaje están en la misma escala de 0 a 100, no es necesario estandarizar.
- Esta técnica se utiliza de manera exploratoria no inferencial.
- La dispersión de los datos, se puede observar en la fase anterior.

Supuestos:

- Representatividad de la muestra. La bondad de los resultados depende directamente de la calidad (representatividad) de los datos considerados. En la siguiente tabla se presenta la cantidad de estudiantes por periodo, distribuidos equitativamente por sexo, los primeros semestres tienden a presentar mejores resultados, dado que representan a los estudiantes de calendario B.

Tabla 10. Número de registros por sexo y medidas por competencia. Año 2014-2 – 2016-2

PERIODO	Sexo	Cantidad	Matemáticas		Lectura	
			Media	Desviación	Media	Desviación
2014-2	Femenino	274.641	49,4	8,8	50,5	9,2
	Masculino	222.039	51,5	9,6	50,1	9,4
2015-1	Femenino	10.649	54,9	14,1	55,6	11
	Masculino	11.129	57,4	14,3	54,4	10,6
2015-2	Femenino	269.350	48,8	10,1	49,8	8,2
	Masculino	219.449	51,9	10,9	49,9	8,3
2016-1	Femenino	7.392	63,2	16,1	59,3	8,3
	Masculino	7.523	66,0	16,7	58,1	8,6
2016-2	Femenino	273.021	49,1	10,4	52,3	9,0
	Masculino	222.472	52,3	10,8	52,4	8,7

Fuente. Elaboración propia

- Multicolinealidad. Las variables que son multicolineales están implícitamente ponderadas con más fuerza. En la fase de preparación se eliminaron las variables con alta correlación.
- Base limpia sin atípicos, en la fase de preparación se realizó la respectiva limpieza y análisis de atípicos.

12.2 Modelo de regresión multinomial

El objetivo de desarrollar esta regresión es encontrar la influencia entre las variables que afecten el desempeño académico de los estudiantes. Dado que la variable dependiente nivel de desempeño, generada a partir del clustering tiene más de dos categorías, y a su vez las variables explicativas son categóricas y continuas no es posible plantear una regresión logística.

Aspectos a tener cuenta:

- Dado que es nominal no existe orden en las categorías, se utilizará la última categoría como respuesta base.

- Se utilizará los coeficientes de los ODDS RATIOS en los que se compara cada tipo de desempeño con el nivel base, es decir se utilizará el nivel bajo de desempeño, con el fin de obtener interpretaciones, tales como, es más probable que un estudiante quede en desempeño alto que bajo si: el nivel educativo de la madre es, paga o no pensión, el tamaño del colegio, etc.
- Se asume que los ODDS RATIOS de cualquier par de categorías es independiente de las demás categorías de respuesta.
- En la vista minable se seleccionaron 28 variables que incluirán en el modelo con 4 grupos de características: (9) aspectos del colegio como jornada de estudio, el carácter, la pensión, entre otros (14) socioeconómicos como estrato, calidad de la vivienda, bienes que posee, entre otros, (3) características propias del estudiante la edad, etnia y género y (2) variables del grupo familiar el ingreso familiar y número personas en el hogar.

Supuestos:

- La variable dependiente no puede ser perfectamente pronosticada a partir de variables independientes para ningún caso.
- Las categorías de la variable dependiente deben ser mutuamente excluyentes y exhaustivas. Para soportar esto, se realizó la técnica de *clustering* la cual debe cumplir esas dos restricciones alta cohesión y bajo acoplamiento.
- La *multicolinealidad* es baja ya que resulta difícil diferenciar entre el impacto de varias variables si no es así. [7]. Dado que el 95% de las variables seleccionadas hasta el momento corresponden a variables categóricas, no es posible calcular el coeficiente de correlación. En estos casos se utiliza la prueba de Chi cuadrado, la cual contrasta si las diferencias observadas entre las dos variables son aleatorias. Se realizaron las pruebas para cada par de variables con el software SPSS el cual genera el p-valor para tomar la siguiente decisión:

Hipótesis nula: Las variables son independientes, con un p-valor menor a 0.05 se rechaza la hipótesis.

Se eliminaron las variables que rechazaron la hipótesis y se dejaron las más representativas para el análisis, por ejemplo, educación de padre y de la madre están altamente relacionadas, se dejó la educación de la madre dado que se ha demostrado que la inteligencia del ser humano es heredada de la madre. [8]

13. Diseño de prueba de los modelos

13.1. Prueba Clustering

El plan de prueba del modelo se basa en:

- Revisar si difieren las medias de cada clúster en cada variable (Centros de los conglomerados finales)
- Revisar si distan entre sí los clústeres es formado lo suficiente (Distancias entre los centros de los conglomerados finales)
- Revisar si hay clúster con pocas observaciones o si difieren mucho de tamaño (Número de casos en cada conglomerado)

13.2. Prueba Regresión Multinomial

En esta técnica será no supervisada por ende no tendrá fase de entrenamiento y prueba. El plan de prueba del modelo se basa en:

- Bondad y ajuste, permite evaluar que tan bien el modelo se ajusta a los datos. Pearson presenta la chi cuadrado. Valores grandes de chi-cuadrado indican un ajuste bajo para el modelo, el p-valor menor a 0.05 indica que el modelo no se ajusta bien a los datos.
- Información de ajuste de los modelos, presenta información sobre si todos los coeficientes del modelo son cero, es decir, si alguno de los coeficientes es estadísticamente significativo, p-valor debe ser menor a 0,05 para concluir que el modelo explica de forma estadísticamente significativa la variable dependiente mejor que el modelo solo con intercepto.
- En esta regresión se cuenta con el Seudo R^2 , es decir la proporción de la varianza que puede explicarse con el modelo. En este caso se utilizó el *Cox y Snell*, se basa en la probabilidad de registro para el modelo en comparación con la probabilidad de registro para un modelo de referencia. Sin embargo, con resultados categóricos, tiene un valor máximo teórico de menos de 1, incluso para un modelo "perfecto". Su uso es simplemente de comparación con los otros modelos.

14. Construcción del modelo

14.1. Modelo Clustering

Los parámetros que se definen para este modelo son:

- El algoritmo que se utilizará corresponde a un clúster k-medias. En este caso ya están definidos la cantidad de clúster esperados: (3) Bajo, medio y alto.
- La distancia utilizada es la euclídea simple
- El número máximo de iteraciones para converger será 10.

- Criterio de convergencia, representa una proporción de la distancia mínima entre los centros iniciales de los clústeres. La iteración cesará si una iteración completa no mueve ninguno de los centros de los clústeres en una distancia superior al uno por ciento de la distancia menor entre cualquiera de los centros.

14.1. Modelo Regresión Multinomial

Los parámetros que se definen para este modelo son:

- Los ODDS-RATIO depende de las unidades en que vengan medidas las variables explicativas. En este sentido, la importancia de cada variable explicativa en el modelo debería medirse por el valor de este resultado. Entre más grande significa más relevante para el modelo.
- Se realizará modelos independientes por competencia, nivel de desempeño lectura y nivel de desempeño matemáticas.
- Se calcularán los Pseudo R cuadrado, Información de ajuste de los modelos Estimación de los parámetros, Pruebas de la razón de verosimilitud
- Se realizaron iteraciones con la combinación de variables que se incluirán en el modelo, más adelante se presentan los resultados de las iteraciones.

15. Evaluar Modelo

15.1. Modelo Clustering

La aplicación de los modelos se está realizando para cada uno de los periodos de análisis, desde 2014-2 y 2016-2, y se desea generar por competencia, es decir, se plantearon dos variables clúster por periodo, uno para matemáticas y otro para lectura. Como se ha mencionado anteriormente, se quiere clasificar a las poblaciones en tres grupos, por ende, no se realizaron pruebas de número de clúster.

En el periodo 2014-2 para la competencia en lectura crítica, el primer centroide fue de 40 nivel bajo de desempeño, el segundo de 52 nivel medio y 64 nivel alto, el último nivel generó mayor número de iteraciones, es decir, la clasificación costo mayor trabajo, en este mismo periodo, en la competencia de matemáticas generó unos centroides con valores levemente mayores con respecto a lectura.

Tabla 11. Características de los cluster. Año 2014-2 – 2016-2

		COMPETENCIAS							
		LECTURA CRITICA				MATEMÁTICAS			
Periodo	Clúster	Centro de clúster	Clasifica en:	Número de iteraciones	Cantidad de registros	Centro de clúster	Clasifica en:	Número de iteraciones	Cantidad de registros
2014-2	1	40	Bajo	3	168.692	53	Medio	5	220.624
	2	52	Medio	4	238.420	42	Bajo	3	204.349
	3	64	Alto	5	89.568	66	Alto	5	71.707
2015-1	1	43	Bajo	3	7.652	43	Bajo	3	9.032
	2	56	Medio	4	8.256	59	Medio	4	8.273
	3	69	Alto	4	5.870	78	Alto	4	4.473

2015-2	1	50	Medio	5	216.228	39	Bajo	3	164.330
	2	41	Bajo	3	147.665	52	Medio	4	233.420
	3	61	Alto	5	124.906	66	Alto	4	91.049
2016-1	1	48	Bajo	3	4.168	47	Bajo	4	4.886
	2	70	Alto	3	3.607	86	Alto	3	4.014
	3	59	Medio	3	7.140	65	Medio	4	6.015
2016-2	1	41	Bajo	3	136.687	63	Alto	4	138.462
	2	53	Medio	3	229.164	51	Medio	4	206.774
	3	64	Alto	3	129.642	38	Bajo	4	150.257

Fuente. Elaboración propia

En los siguientes tres gráficos se presenta la dispersión de los puntajes en lectura crítica vs matemáticas marcando en colores los tres conglomerados generados. Adicionalmente, se muestra cada gráfico de dispersión clasificado en el nivel educativo predominante de los profesores de los colegios y finalmente, se presenta la distribución para el departamento del Cauca, Cundinamarca y Bogotá. En estos gráficos es posible identificar:

- Una buena clasificación del clustering, mostrando en color azul claro el nivel bajo, seguido del naranja en nivel medio y azul oscuro en nivel alto.
- En el gráfico 23 se presenta la distribución del departamento de Cauca, donde se resalta la cantidad de colegios que predominan los profesores bachilleres.
- A diferencia del Bogotá, que no registran colegios que predominen bachilleres, en la capital se concentran en profesores de pregrado no licenciados.
- Finalmente, en el caso de Cundinamarca, registra mayor número de colegios con nivel alto de profesores con posgrado, seguramente eso se debe a los colegios más grande de Bogotá que están ubicados a las afueras de la ciudad.

Gráfico 23. Dispersión de los puntajes lectura vs matemáticas por categoría profesores en CAUCA. 2015-2

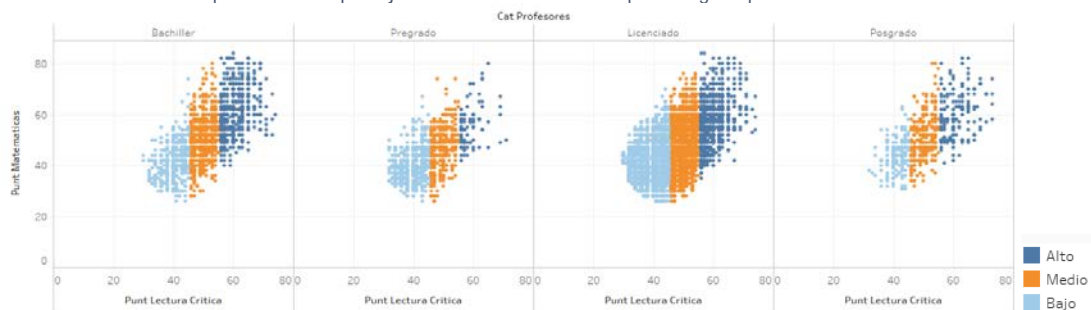


Gráfico 24. Dispersión de los puntajes lectura vs matemáticas por categoría profesores en BOGOTÁ. 2015-2

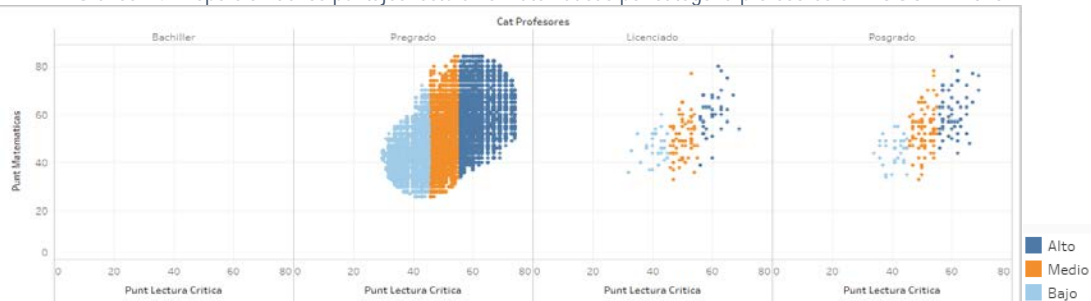
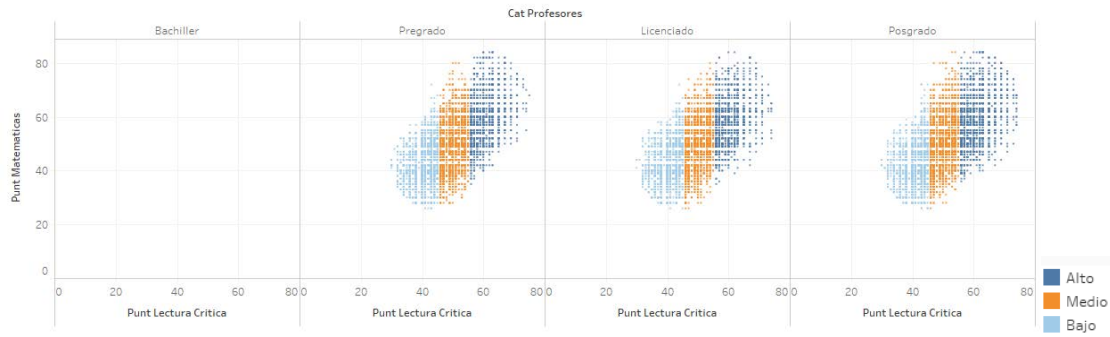


Gráfico 25. Dispersión de los puntajes lectura vs matemáticas por categoría profesores en CUNDINAMARCA 2015-2



Fuente. Elaboración propia

Lo anterior permite deducir que el resultado de la aplicación del modelo tiene sentido en el contexto de educación y resultó una buena generación de subgrupos, los cuáles serán explicados a mayor detalle con la ayuda de la regresión multinomial.

15.2. Modelo Regresión Multinomial

Para la regresión multinomial se corrieron varios modelos con variable dependiente el clúster generado anteriormente, es decir, nivel de desempeño y la combinación de las siguientes variables independientes, seleccionadas a partir del análisis exploratorio realizado en la fase 2 y adicionalmente, tomando como buenas prácticas los trabajos de investigación realizados a la misma fuente de información, tal como el artículo sobre Brechas de género [5]. En la iteración se encontró algo que llama mucho la atención, puesto que los resultados de los modelos iban mejorando en la medida que se aumentaban las variables referentes al colegio.

Tabla 12. Iteraciones de la regresión multinomial

Variables	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
Edad simple				X	
Edad categórica	X	X	X		X
Género	X	X	X		
Pertenece etnia	X			X	
Zona de residencia		X			X
¿Paga pensión?	X		X	X	
Educación madre	X	X	X	X	X
Ocupación padre	X			X	X
Ocupación madre		X	X		
Estrato vivienda	X	X	X	X	X
Personas en el hogar		X			X
Ingreso familiar	X			X	
Tiene lavadora			X		
Tiene pc	X			X	
Género colegio			X		
Naturaleza colegio	X	X	X	X	X
Jornada colegio	X	X	X	X	
Calendario colegio			X	X	
Cantidad estudiantes		X	X		
Categoría colegio	X		X		X
Nivel educativo profesores			X		X

Fuente. Elaboración propia

Cómo se mencionó en el apartado de plan de prueba de los modelos, se tuvieron en cuenta 4 estadísticos para tomar la decisión del “mejor” modelo con las variables disponibles, el cual logrará explicar en mejor medida la relación de las variables con el desempeño académico de los estudiantes.

Con relación a la bondad y ajuste se presentan dos medidas que evalúan que tan bien el modelo se ajusta a los datos, valores muy grandes del Chi cuadrado indican un ajuste pobre del modelo y un valor estadísticamente significativo es decir $p < 0,5$ indica que el modelo no se ajusta bien a los datos. En este sentido, los modelos 1, 4 y 5 no cumplen con estos valores por lo cual se deben rechazar.

La tercera línea responde si al menos una variable incluida en el modelo estadísticamente de manera significativa mejora el modelo en comparación con el intercepto solo, para cumplir esto el p -valor $< 0,05$. En este caso todos los modelos cumplieron con esta validación.

Tabla 13. Iteraciones de la regresión multinomial

Parámetros de evaluación	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
Bondad y ajuste, Chi-Cuadrado	136.296	96.260	94.143	689.606	116.334
P-valor del Chi-Cuadrado	0	0,143	0,361	0	0
P-valor del intercepto	0	0	0	0	0
Cox y Snell	0,245	0,53	0,58	0,14	0,35

Fuente. Elaboración propia

Finalmente, se decidió utilizar la combinación de variables del modelo 3, el cual cumple en mejor medida con las pruebas de evaluación.

FASE 5. EVALUACIÓN

En esta fase, el modelo se evalúa exhaustivamente para asegurarse de que logra los objetivos comerciales, se revisa todo el proceso y se determinan los siguientes pasos.

16. Evaluar los resultados

Esta podría ser la más importante más relevante dentro del desarrollo de la metodología, puesto que en esta sección se condensarán los resultados de los modelos con el análisis y objetivo del negocio, en esta sección cobra sentido los resultados del mismo y se identifican los aportes de generación de valor para el negocio, es decir, brindar solución al problema o brindar nueva información con la que no contaban.

En este proyecto particularmente, el problema se basa en la búsqueda de información relevante a partir de las fuentes disponibles y accesibles que estén relacionados con el objetivo de educación de calidad de los ODS. Tal como se ha indicado anteriormente, en este primer ejercicio se profundizó en la *Meta 6. Alfabetización y conocimiento elementales de aritmética en jóvenes*, actualmente es posible tener un acercamiento a esta meta a partir de las pruebas saber 11, donde mide el nivel de conocimiento en las competencias de lectura crítica y matemáticas.

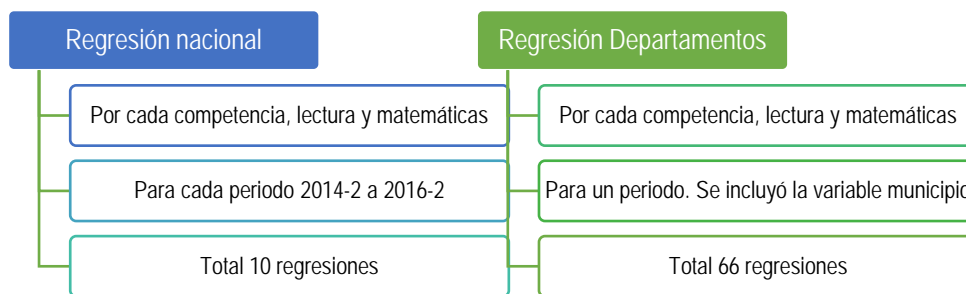
En la tabla 13 se presentan el porcentaje de estudiantes por nivel de desempeño en las competencias, lectura crítica y matemáticas, se logra ver que el porcentaje de estudiantes en el desempeño bajo está decreciendo siendo esto un buen indicador para la educación en Colombia. No obstante, este porcentaje no genera el conocimiento sobre las variables que afectan para que esto suceda o si existen diferencias entre regiones del país.

Tabla 14. Porcentaje de estudiantes por nivel de desempeño por competencia. Periodo 2014-2 a 2016-2

Competencia	Nivel	2014-2	2015-2	2016-2
Lectura	Bajo	34,0%	30,2%	27,6%
	Medio	48,0%	44,2%	46,2%
	Alto	18,0%	25,6%	26,2%
Matemáticas	Bajo	41,1%	33,6%	30,3%
	Medio	44,4%	47,8%	41,7%
	Alto	14,4%	18,6%	27,9%

Fuente. Resultados saber 11. ICFES

En este sentido, se plantea la regresión multinomial para que permita obtener mayor información sobre estos resultados. Se han planteado las siguientes desagregaciones que permitan comparar resultados durante el periodo de análisis y desagregado por departamento y municipio.



Se corrió una regresión para cada periodo con la variable dependiente nivel de desempeño y definiendo como categoría base, nivel bajo de desempeño, es decir, se analizará qué variables afectan para incrementar la probabilidad de ser nivel medio o nivel alto de desempeño con base en la probabilidad de desempeño bajo

Dentro de las medidas que se genera en spss a partir de la función NOMREG, se deben tener en cuenta:

- En la tabla de pruebas de la razón de verosimilitud, se debe verificar que cuáles variables son significativas para el modelo a partir de la columna SIG., donde los valores menores a 0,05 indican que si son significantes.
- En la tabla de estimaciones de parámetros, se deben analizar los resultados en dos conjuntos, por una parte, una comparación entre los estudiantes de desempeño bajo con desempeño medio y otro conjunto de los estudiantes de desempeño bajo con desempeño alto. Dentro de esta tabla, se deben revisar tres medidas: el valor del Exp(B) mismo Odds Ratio, entre más grande mejor, los intervalos del Exp(B), si este intervalo el valor 1, se concluye que la asociación no estadísticamente significativa y finalmente si el intervalo no incluye en 1, se revisa el signo del coeficiente (B).

Resultados

Para ilustrar la manera de analizar los resultados de las tablas, a continuación, se muestran los resultados de la regresión para el periodo 2015-2 con la variable dependiente desempeño en **matemáticas**. La tabla 15 permite identificar las variables significativas en el modelo, en este caso la única variable que no resultó significativa fue calendario del colegio.

Al comparar estos resultados por periodo, se identificó que este resultado es igual para los segundos semestres, es decir en los 2014-2, 2015-2 y 2016-2 la variable no significativa es calendario del colegio. Sin embargo, en los primeros periodos, aumentan las variables no significativas, tales como, el tamaño del colegio en términos de estudiantes, si paga o no pensión, la cantidad de personas en el hogar y en cambio el calendario si es significativa. Finalmente, estos resultados no tienen importantes diferencias por competencia.

Tabla 15. Pruebas de la razón de verosimilitud de la primera regresión

Efecto	Pruebas de la razón de verosimilitud		
	Chi-cuadrado	gl	Sig.
Intersección	0,000	0	
EVALUADOS_ULTIMOS_3	133,690	2	,000

ESTU_ETNIA	1179,534	2	,000
Edad_c	8129,745	2	0,000
CAT_PROFESORES	1171,396	8	,000
CATEGORIA_COLEGIO	22123,718	8	0,000
COMPUTADOR	811,667	2	,000
LAVADORA	56,153	2	,000
NATURALEZA_C	70,144	2	,000
JORNADA_CATEGORIA	4629,745	2	0,000
CALENDARIO_C	1,417	2	,492
GENERO_C	99,843	4	,000
GENERO	12906,868	2	0,000
FAMI_PERSONAS_HOGAR	247,191	4	,000
FAMI_ESTRATO_VIVIENDA	1063,040	10	,000
FAMI_EDUCA_MADRE	3212,652	12	0,000
FAMI_OCUPA_MADRE	234,229	16	,000
COLE_VALOR_PENSION	40,458	2	,000

Fuente. Elaboración propia

Ahora bien, en la tabla 16 se presentan las medidas que permiten interpretar el modelo, como se explicó anteriormente, luego de confirmar que el intervalo de confianza no contenga el 1, se puede interpretar el resultado del Odd Ratio Exp(B).

Para el periodo 2015-2 se identificó para las columnas de color verde que:

1. Es más probable que un estudiante quede en desempeño medio que bajo si cuando presenta el examen es menor de edad.
2. Es más probable que un estudiante quede en desempeño medio que bajo si el nivel educativo predominante de los profesores es posgrado o licenciado.
3. Es más probable que un estudiante quede en desempeño medio que bajo si la clasificación del colegio es alta. Particularmente esta variable tiene mayor número del Odd Ratio y las demás están en valores muy cercanos entre ellas. Más adelante se presentarán conclusiones al respecto.
4. Es más probable que un estudiante quede en desempeño medio que bajo si estudia en colegio privado y en jornada diurna
5. Es más probable que un estudiante quede en desempeño medio que bajo si la madre tiene un nivel educativo mayor a primaria y tiene alguna actividad diferente al hogar.
6. Es más probable que un estudiante quede en desempeño medio que bajo si convive con pocas personas en el hogar.

Para el periodo 2015-2 se identificó para las columnas de color naranja que:

7. Es más probable que un estudiante quede en desempeño alto que bajo si el nivel educativo predominante de los profesores es posgrado.
8. En relación a las demás variables los resultados fueron los mismos, sin embargo, hay variables que aumentan la probabilidad mucho más de ser tener un desempeño alto, por ejemplo, la categoría del colegio es fundamental, que la madre tenga específicamente posgrado y que la jornada del colegio no sea nocturna.

Al comparar estos resultados por periodo, se identificó que este resultado es igual para todos los semestres, tan solo en los primeros semestres se comienza generar importancia de la variable si paga o no pensión, se debe revisar la calidad de esta variable pues se registran colegios de calendario B,

es decir privados que no pagan pensión. Finalmente, estos resultados no tienen importantes diferencias por competencia.

Tabla 16. Estimación de parámetros período 2015-2

Número de caso de clúster ^a	Desempeño Medio				Desempeño Alto			
	B	Exp(B)	95% de intervalo de confianza para Exp(B)		B	Exp(B)	95% de intervalo de confianza para Exp(B)	
			Límite inferior	Límite superior			Límite inferior	Límite superior
[Edad_c=1,00]	,515	1,674	1,648	1,700	1,058	2,881	2,808	2,956
[CAT_PROFESORES=2]	,297	1,346	1,265	1,433	,366	1,443	1,332	1,562
[CAT_PROFESORES=4]	,135	1,145	1,076	1,218				
[CATEGORIA_COLEGIO=3]	1,739	5,692	5,311	6,100	3,526	33,973	31,589	36,536
[NATURALEZA_C=2]	,139	1,150	1,078	1,226	,357	1,430	1,315	1,554
[JORNADA_CATEGORIA=1,00]	,578	1,782	1,743	1,822	1,319	3,741	3,554	3,938
[FAMI_PERSONAS_HOGAR=2,00]	,180	1,197	1,152	1,243	,332	1,394	1,311	1,481
[FAMI_EDUCA_MADRE=10]	,582	1,789	1,597	2,003	1,460	4,307	3,775	4,916
[FAMI_OCUPA_MADRE=2]	,111	1,118	1,063	1,175	,223	1,250	1,175	1,329

Fuente. Elaboración propia

Ahora bien, se analizaron los departamentos que cubren el 50% de la población total: Bogotá, Antioquia, Valle del Cauca, Cundinamarca.

A las regresiones que se aplicaron a los departamentos se incluyó la variable municipio, con el fin de identificar si existen municipios que influyan más en el desempeño general del departamento, a continuación, se presenta algunas ideas principales de este análisis departamental.

- Antioquia, curiosamente define la naturaleza del colegio como una variable no significativa, es decir si es público o privado. La distribución es similar a la nacional, 75% públicos. El modelo identifica tres municipios que aumentan la probabilidad de tener desempeño medio con respecto al bajo y particularmente se evidencia 5 municipios con ODD Ratios superior a 3, es decir que aumentar la probabilidad de que se categorice en desempeño alto.
- Bogotá, no identifica significativa los bienes del hogar y los resultados son altamente parecidos a los nacionales.
- Atlántico, esta regresión no identifica significativa la pertenencia a etnia, la ocupación de la madre y si paga o no pensión. Unas de las variables diferentes que este modelo genera importantes es el sexo del colegio, es decir si es femenino aumenta la probabilidad de quedar en desempeño medio que en bajo. Ningún municipio generó efecto importante.
- Valle del Cauca, no es significativa el tamaño del colegio, las interpretaciones de las variables se asemejan a las nacionales y no hay diferencias importantes entre municipios.
- Cundinamarca, en este departamento el nivel educativo de los profesores es importante y no se registran diferencias importantes por municipio.

En definitiva, luego de los análisis realizados es posible concluir que el colegio y la calidad educativa del mismo hace un papel determinante en los resultados académicos de los estudiantes, así mismo el

nivel educativo de la madre y la edad de presentación del examen son las variables que permanecen significativas y de valor en todas las iteraciones del modelo.

En este sentido, se desea profundizar cuáles son las características de un colegio clasificado en el mejor rango, como se presentó en el gráfico 18, la distribución de los colegios los colegios A+, es decir con los mejores puntajes, el 89,37% corresponde a colegios privados y el 10,63% a colegios públicos, y cabe aclarar que el 76,3% de los estudiantes que presentaron la prueba son de colegios públicos, esto indica que a los colegios con alta calidad de educativa sólo acceden muy pocos estudiantes.

En los siguientes gráficos se presenta algunas características de los colegios A+, las cuales permitirán generar recomendaciones para la mejora en la calidad educativa, por una parte, es evidente que el nivel educativo de los docentes afecta directamente los resultados, más del 70% de los colegios A+ se caracterizan por tener profesores con posgrado.

Revisando los resultados de las diferentes pruebas del examen, inglés, ciencias naturales, ciencias sociales, lectura y matemáticas, se evidencia que los colegios A+ la competencia inglés es la que presenta mejores puntajes, seguido de matemáticas, sin embargo, en las demás categorías de colegios esta característica en inglés se pierde, finalmente, si hay una evidente diferencia en el nivel educativo de la madre en los colegios A+, donde se registran gran parte de las madres con posgrado y el nivel de profesional es el predominante.

En las recomendaciones finales, se relacionan algunas ideas relacionadas con la posibilidad de plantear acciones que afecten las variables importantes y que están en manos del gobierno a través de políticas públicas.

Gráfico 26. Distribución de los colegios clasificados según el icfes y el nivel educativo de los profesores 2015-2

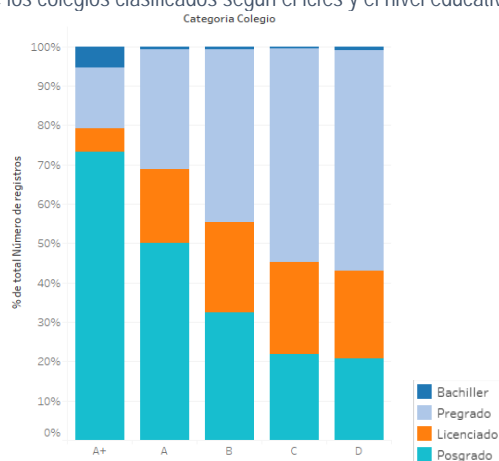


Gráfico 27. Distribución de los colegios clasificados según el icfes y el puntaje promedio de las competencias. 2015-2

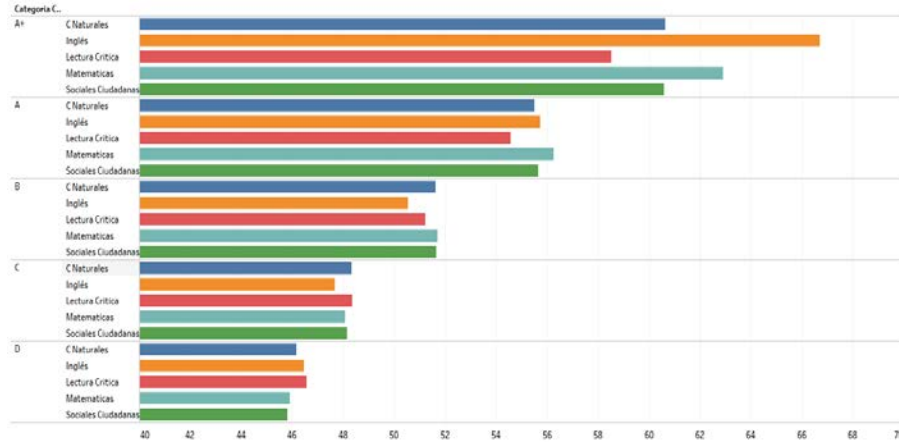
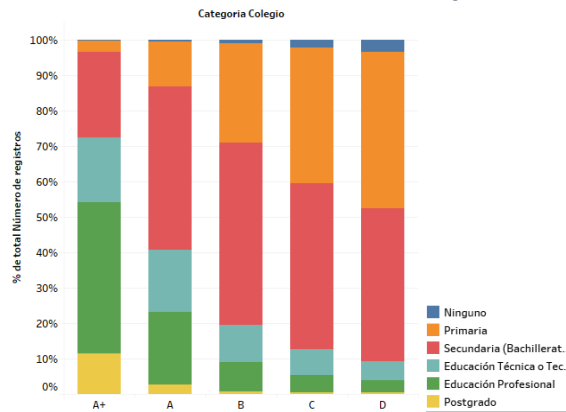


Gráfico 28. Distribución del educativo de la madre según categoría del colegio. 2015-2



17. Revisión del proceso

Al iniciar el proceso lo primero que se requirió fue definir la herramienta de integración, con las capacidades de tolerar el volumen de datos, rapidez en el proceso y bajos consumo de recursos de hardware, este proceso requirió pruebas de configuración de varias herramientas y de desempeño teniendo que hacer configuraciones similares en cada herramienta repitiendo el proceso de carga o complementando los avances obtenidos en otras herramientas, esto con llevó a un aumento en los tiempos para esta tarea específica. Dado que esto hace parte importante del desarrollo del proyecto hacer configuraciones que permitan repetir los procesos de unificación de datos en caso de tener que repetirlos de tal forma que solo se invierta en los tiempos de procesamiento y una configuración mínima.

Dado la naturaleza del proyecto, las fuentes de datos eran diversas haciendo que la unificación de formatos y valores fuera extensa y compleja, adicionalmente a medida que se encontraba más información, era necesario replantear o repetir el proceso y hasta regresar a verificar los valores ya unificados para ajustarse a los nuevos datos. En algunos casos se evidenció que, siendo información

del mismo sector gubernamental, y que deberían contar con descripciones unificadas y codificación estandarizada, no tienen esta cultura implicando cambio de nombres y re categorizaciones.

En el caso del proyecto se realizó un proceso de unificación de datos por la temática que compartían, saber11; luego de un proceso descriptivo y detallado de los datos unificados, se llegó a la conclusión de restringir los datos a periodos puntuales o archivos especiales, lo que se resume en pérdidas de esfuerzo en la unificación, por esto se considera valioso hacer una revisión independiente de los datos para luego identificar que se debe consolidar y que no enfocando de mejor manera los esfuerzos.

18. Determinar los próximos pasos

En la revisión de fuentes disponibles se identificaron algunas con las cuales se puede completar la base de datos trabajada en este proyecto, tales como:

- Observar la relación la tasa de cobertura, tasa de matrícula, tasa de deserción por municipio. Fuente del Ministerio de Educación Nacional, disponible en la página web y actualmente, almacenada en el prototipo del observatorio.
- Se cuenta con una fuente proporcionada por el Ministerio de las TIC's, con variables interesantes por municipio, dentro del proyecto "computadores para educar" tabletas y portátiles entregados, así como profesores capacitados en TIC's.
- Otro análisis interesante es revisar, los movimientos de lugares de residencia, lugar de estudio y lugar de presentación del examen, sobre todo en municipios y de esta manera tener un acercamiento a movilidad interna por estudio.

Anteriormente, se mencionó que la base de datos se filtró con los estudiantes que se presentaron por parte de un colegio, dado que necesitábamos completitud en estas variables, sin embargo, aún queda la posibilidad de analizar brechas entre estas dos poblaciones para proponer estrategias y disminuir esas diferencias que actualmente se evidencian.

Siguen como alternativas, las propuestas analíticas relacionados en la fase del entendimiento del negocio:

- Influencia de las competencias TIC de los egresados de pregrado en los ingresos después de año de grado.
- Valor agregado de las instituciones de educación superior dependiendo del colegio de procedencia

MANEJO RESPONSABLE DE LA INFORMACIÓN

En Colombia hubo un boom con el manejo responsable de la información cuando se publicó la ley 1581 del 2012, que generó una conciencia clara de la importancia de la información personal que todas las instituciones manejan de sus clientes y proveedores; así mismo se hizo claro que toda la información debe ser tratada con lineamientos claros de seguridad e integridad para que esta sea resguardada y restringida.

Para la Alianza Caoba el manejo responsable de la información, es uno de los pilares sobre los que se fundamenta, desde la concepción de la Alianza que propone como uno de sus intereses "Promover cultura de Open Data" y "Acceder a datos reales" entre otros intereses, más los preceptos institucionales que conlleva la responsabilidad que se tiene al tener acceso a información de las organizaciones que se analizará en los proyectos de Big Data y Analytics.

Para contextualizar el término Open Data se hace referencia a: datos que pueden ser utilizados libremente, reutilizado y redistribuidos, según la referencia se caracterizan por:

Disponibilidad y acceso: los datos deben estar disponibles en su totalidad y no deben exceder un costo de reproducción razonable, preferiblemente mediante la descarga a través de Internet.

Reutilización y redistribución: los datos deben proporcionarse en términos que permitan su reutilización y redistribución, incluida la combinación con otros conjuntos de datos.

Participación universal: todos deben poder utilizar, reutilizar y redistribuir, no debe haber discriminación en los campos de actividad o en contra de personas o grupos.

Retomando el planteamiento del proyecto, este se propone para que utilice como fuente de información aquella que sea pública y abierta (Open Data), por esto desde la concepción y consecución de los datos, el proyecto cuenta con los lineamientos de manejo responsable, como son: la anonimización, la autorización para ser publicada y en consecuencia consultada y utilizada para la investigación.

CONCLUSIONES Y RECOMENDACIONES

A partir de los objetivos del proyecto se relacionarán las conclusiones y recomendaciones:

En la relación a la caracterización de los indicadores de las 7 metas del objetivo de educación de calidad, se convirtió en un producto importante para el inicio del desarrollo del proyecto, dado que allí se logró identificar un número importantes de fuentes que actualmente están disponibles para análisis y generación de valor en temas de educación de calidad en Colombia. Para definir las fuentes que se presentaron en el producto de caracterización, se tuvieron en cuenta las fuentes que propone la UNESCO y las fuentes que por conocimiento en el tema educativo se lograron obtener, no obstante, esto no significa que efectivamente sean las fuentes oficiales utilizadas por Colombia para responder a estos indicadores, incluso los indicadores propuestos por la UNESCO, son precisamente propuestas para que los países que permitan comparaciones, pero estos indicadores pueden presentar ajustes por expertos nacionales en los casos en cuales Colombia no defina o calcule de la misma manera los indicadores.

En este sentido, se recomienda a Alianza Caoba realizar entrevistas con personas encargadas de los Objetivo de Desarrollo Sostenible en las entidades oficiales, para corroborar los indicadores adoptados para Colombia y la confirmación de las fuentes relacionadas. A continuación, se relacionan los datos de los contactos encargados de lo ODS:

- Helga Milena Hernández, Jefe Oficina Asesora de Planeación y Finanzas
Ministerio de Educación Nacional
Secretaría: Gloria Cañón. Teléfono: 489 0400 Ext:1509
Correo. planeacion@mineducacion.gov.co
- Diana Carolina Nova Laverde, Líder de los Objetivos de Desarrollo Sostenible de la Dirección General.
Departamento Administrativo Nacional de Estadística
Teléfono: 5978300 Ext: 3209
Correo. dcnoval@dane.gov.co

En la construcción del gobierno de datos, uno de los inconvenientes que se identificaron al hacer el proceso descriptivo y el análisis de los datos, fueron las falencias en la calidad de la información que impactaron los tiempos y que hicieron explícita la usencia y la importancia de la definición y divulgación de estándares para el manejo de la misma, esta ausencia hizo notorio y valioso el que este proyecto incluya un documento gobierno de datos, que de los lineamientos para el manejo de estos y defina las categorías válidas y sus significados, permitiendo así que toda la información esté articulada y que la nueva se organice basada en las políticas definidas o bien que den paso a nuevas políticas que incluyan aspectos que no se han tenido en cuenta y que den paso a los nuevo datos.

Luego de conocer los datos y la información que se puede extraer de estos, se ve la necesidad de poder visulizarla de una manera comprensible, que no requiera conocimientos avanzados y por ende

que sea del dominio público, este es el caso del observatorio que se propone, puesto que facilita la comprensión y la difusión a la mayoría de personas de una manera ágil y fácil de entender, así las cosas la prueba prototipo habilita este mecanismo desde internet y con una estructura gráfica sencilla, didáctica y flexible, para que el usuario final interactúe con los datos, permitiendo así visualizar y profundizar en los aspectos de su interés.

Finalmente, con respecto al análisis realizado a la fuente de resultados del examen de estado Saber 11, se lograron identificar variables importantes que tienen influencia en los puntajes de las competencias en lectura crítica y matemáticas. Luego generar regresiones multinomiales por periodo, competencias y departamentos, se llegó a la conclusión que el nivel de formación de los docentes de los colegios tienen efecto en los puntajes, los colegios con mejores resultados tiene una marcada diferencia en la segunda lengua, el nivel educativo de las madres, específicamente posgrado aumenta la probabilidad de que un estudiante tenga desempeño alto y finalmente, la edad del estudiante en el momento de presentar el examen genera mejores puntajes.

De acuerdo con lo anterior, las recomendaciones se resumen así:

- Incentivar sobre todo en los colegios públicos que representa más del 70% del país, el ascenso en el desarrollo profesoral a través de la formación docente. Adicionalmente, propender por estudios de posgrado del área educativa.
- Fortalecer en todos los colegios, la formación bilingüe, incluir literatura minina en inglés al año, La importancia de la variable edad, referente a mejores resultados siendo menor de edad se puede interpretar como los efectos en los estudiantes que han desertado o aplazados sus estudios de bachillerato, según los datos abiertos del MEN⁹ entre el 5 y 10% de los estudiantes desertan de la educación secundaria, en este sentido, aparte de focalizar esfuerzos en cobertura, se deben analizar los aspectos de la deserción a este nivel.

⁹ <https://www.datos.gov.co/Educaci-n/ESTADISTICAS-EN-EDUCACI-N-BASICA-POR-ETC/sras-4t5p/data>

BIBLIOGRAFÍA

- [1] «Naciones Unidas, CEPAL (2016). Agenda 2030 y los Objetivos de Desarrollo Sostenible Una oportunidad para América Latina y el Caribe. Santiago de Chile». .
- [2] «Metadata_for_SDG_4_Global_and_thematic_indicators_oct.pdf». .
- [3] «Decreto 869 de 2010. Reglamenta saber11». .
- [4] R. J. Murnane, J. B. Willett, Y. Duhaldeborde, y J. H. Tyler, «How Important Are the Cognitive Skills of Teenagers in Predicting Subsequent Earnings?», *J. Policy Anal. Manage.*, vol. 19, n.º 4, pp. 547-568, 2000.
- [5] L. K. A.-G. Bernal y G. Bernal, «Brechas de género en el rendimiento escolar a lo largo de la distribución de puntajes: evidencia pruebas saber 11°», UNIVERSIDAD JAVERIANA - BOGOTÁ, 015301, sep. 2016.
- [6] «Bol_nbi_censo_2005.pdf». .
- [7] E. H. Slate, «Conditioning Diagnostics: Collinearity and Weak Data in Regression», *J. Am. Stat. Assoc.*, vol. 88, n.º 421, pp. 384–386, 1993.
- [8] G. Turner y M. W. Partington, «Genes for intelligence on the X chromosome.», *J. Med. Genet.*, vol. 28, n.º 6, p. 429, jun. 1991.

ANEXOS

ANEXO 1. Fuentes datos disponibles

Datos Abiertos Colombia (datos.gov.co) *

- Número de visitas a bibliotecas públicas
- Relación Instituciones Educativas Beneficiadas Del Plan De Alimentación Escolar (Pae)

ICFES - Pruebas de estado:

- Resultados de las pruebas Saber 11 desde el año 2000 hasta 2017
- Resultados de las pruebas Saber PRO desde el año 2011 hasta 2016
- Resultados de las pruebas Saber 5 y 9 desde el año 2002 hasta el 2009
- Base de datos que permite relacionar aproximadamente 200.000 estudiantes de las bases saber 11 y saber pro.

SNIES, Educación Superior:

- Información por institución de educación superior: inscritos, admitidos, primer curso, matriculados, graduados, docentes y administrativos 2013-2016 (Agregados)

SPADIES, Educación Superior: herramienta para hacer seguimiento a la deserción del país

Ministerio de Educación Nacional

- Observatorio laboral para la educación, información sobre ingreso y tasa de cotización
- Establecimientos educativos por Municipio *
- Establecimientos educativos de preescolar, básica y media*
- Estadísticas en educación de preescolar, básica y media por entidad territorial*
- Estadísticas en educación de preescolar, básica y media por municipio*
- Estadísticas en educación superior por municipio*
- Instituciones de educación de formación para el trabajo y el desarrollo humano*

Ministerio de Trabajo, sistema estadístico de información laboral formal

Departamento Administrativo Nacional de Estadística

- Acceso a Internet 10-14 (registro)
- Gran encuesta integrada de hogares (fuerza laboral y educación)
- Encuesta de Calidad de Vida (2003 – 2016)
- Censos Generales 2005, 1993, 1985
- Proyecciones de población hasta 2020 realizadas a partir del censo 2005
- Encuesta de convivencia escolar Bogotá
- Encuesta de comportamientos y actitudes sexuales en población escolarizada
- Encuesta de formación de capital humana
- Encuesta para formación para el trabajo
- Índice de costos de la Educación Superior (registros IES)

Sistema de Naciones Unidas

- Base de datos sobre la desigualdad en la educación
- Bases de datos a nivel mundial que responden a los indicadores del cuarto objetivo de desarrollo sostenible.

Estudios Mundiales, resultados pruebas TERCE

ANEXO 2. Información de metadatos

Proceso de Extracción		
Descripción General de las fuentes de datos	Archivos csv que contienen la información semestre a semestre de las personas que han presentado el examen saber 11 desde el año 2000 primer semestre y el año 2017 primer semestre.	
Descripción General del proceso	Limpieza de datos para el procesamiento analítico.	
Describir si hay relación entre ellos detalladamente	Los archivos en su mayoría comparten las mismas columnas entre sí pero en algunos periodos de tiempo se van incrementando, cuentan con información demoGráfico resultados del examen e información de interés general.	
Listado de productos de Software utilizados en el proceso		
Nombre del Software	Tipo de Licenciamiento	Versión
SPSS	Institucional educativa	24
RapidMiner	Educativa	
Tableau	Temporal public	
Excel	Enterprise	2016

SQL Server	Educativa	2016
Información sobre los Datasets		
Nombre del Archivo	Tamaño en Bytes o MD5	Formato
SB11-20001-RGSTRO-CLFCCN-V1-0.txt	17.948.011	txt
SB11-20002-RGSTRO-CLFCCN-V1-0.txt	88.015.028	txt
SB11-20011-RGSTRO-CLFCCN-V1-0.txt	18.149.265	txt
SB11-20012-RGSTRO-CLFCCN-V1-0.txt	78.712.147	txt
SB11-20021-RGSTRO-CLFCCN-V1-0.txt	15.035.745	txt
SB11-20022-RGSTRO-CLFCCN-V1-0.txt	91.320.606	txt
SB11-20031-RGSTRO-CLFCCN-V1-0.txt	16.136.443	txt
SB11-20032-RGSTRO-CLFCCN-V1-0.txt	87.574.158	txt
SB11-20041-RGSTRO-CLFCCN-V1-0.txt	15.831.241	txt
SB11-20042-RGSTRO-CLFCCN-V1-0.txt	87.309.892	txt
SB11-20051-RGSTRO-CLFCCN-V1-0.txt	30.532.289	txt
SB11-20052-RGSTRO-CLFCCN-V1-0.txt	163.889.480	txt
SB11-20062-RGSTRO-CLFCCN-V1-0.txt	176.402.858	txt
SB11-20071-RGSTRO-CLFCCN-V1-0.txt	28.096.282	txt
SB11-20072-RGSTRO-CLFCCN-V1-0.txt	195.099.386	txt
SB11-20081-RGSTRO-CLFCCN-V1-0.txt	27.241.230	txt
SB11-20082-RGSTRO-CLFCCN-V1-0.txt	184.704.179	txt
SB11-20091-RGSTRO-CLFCCN-V1-0.txt	32.747.328	txt
SB11-20092-RGSTRO-CLFCCN-V1-0.txt	241.778.388	txt
SB11-20101-RGSTRO-CLFCCN-V1-0.txt	12.407.460	txt
SB11-20101-RGSTRO-CLFCCN-V1-0.txt	12.407.460	txt
SB11-20102-RGSTRO-CLFCCN-V1-0.txt	242.150.147	txt
SB11-20111-RGSTRO-CLFCCN-V1-0.txt	35.829.811	txt
SB11-20112-RGSTRO-CLFCCN-V1-0.txt	222.152.209	txt
SB11-20122-RGSTRO-CLFCCN-V1-0.txt	229.759.019	txt
SB11-20131-RGSTRO-CLFCCN-V1-0.txt	31.255.400	txt
SB11-20132-RGSTRO-CLFCCN-V1-0.txt	217.961.121	txt
SB11-20141-RGSTRO-CLFCCN-V1-0.txt	28.103.011	txt
SB11-20142-RGSTRO-CLFCCN-V3-0.txt	239.117.583	txt
SB11-20151-RGSTRO-CLFCCN-V3-0.txt	38.492.014	txt
SB11-20152-RGSTRO-CLFCCN-V3-0.txt	210.154.050	txt
SB11-20161-RGSTRO-CLFCCN-V3-0.txt	20.711.008	txt
SB11-20162-RGSTRO-CLFCCN-V1-0.txt	199.341.386	txt
SB11-20171-RGSTRO-CLFCCN-V1-0.TXT	10.379.638	txt

ANEXO 3. Resumen de variables. Final del documento.

CENTRO DE EXCELENCIA Y APROPIACIÓN BIG DATA Y DATA ANALYTICS- CAOBA



OBJETIVO DE DESARROLLO SOSTENIBLE:
EDUCACIÓN DE CALIDAD

CARACTERIZACIÓN DE INDICADORES

BOGOTÁ, D.C., NOVIEMBRE 2017

Contenido

CARACTERIZACIÓN DE INDICADORES.....	61
Meta 1. Enseñanza primaria y secundaria	61
Indicador Mundial 1.1.....	61
Indicador Temático 1.2.....	63
Indicador Temático 1.3	64
Indicador Temático 1.4.....	65
Indicador Temático 1.5.....	66
Indicador Temático 1.6.....	70
Indicador Temático 1.7.....	71
Meta 2. Primera infancia.....	73
Indicador Mundial 2.1	73
Indicador Mundial 2.2.....	75
Indicador Temático 2.3.....	78
Indicador temático 2.4.....	80
Indicador temático 2.5.....	81
Meta 3. Enseñanza técnica, profesional, superior y de adultos	83
Indicador Mundial 3.1.....	83
Indicador Temática 3.2.....	85
Indicador Temática 3.3.....	87
Meta 4. Competencias para el trabajo.....	89
Indicador Mundial 4.1.....	89
Indicador Temático 4.2.....	91
Indicador Temático 4.3.....	93
Meta 5. Equidad	95
Indicador Mundial 5.1.....	95
Indicador Temático 5.2.....	97
Indicador Temático 5.3.....	99
Indicador Temático 5.4.....	101
Indicador Temático 5.5.....	103
Meta 6. Alfabetización y conocimientos elementales de aritmética.....	103
Indicador Mundial 6.1	104
Indicador Temático 6.2.....	105

Indicador Temático 6.3.....	108
Meta 7. Desarrollo sostenible y ciudadanía mundial	110
Indicador Mundial 7.1	110
Indicador Temático 7.2.....	111
Indicador Temático 7.3.....	113
Indicador Temático 7.4.....	114
Indicador Temático 7.5.....	116
Meta 7A. Instalaciones escolares y entornos de aprendizaje	117
Indicador Mundial 7A.1	117
Indicador Temático 7A.2	119
Indicador Temático 7A.3	121
Meta 7B. Becas	121
Indicador Mundial 7B.1	122
Indicador Temático 7B.2	122
Meta 7C. Docentes.....	123
Indicador Mundial 7C.1	123
Indicador Temático 7C.2	124
Indicador Temático 7C.3	126
Indicador Temático 7C.4	128
Indicador Temático 7C.5	130
Indicador Temático 7C.6	131
Indicador Temático 7C.7	133

CARACTERIZACIÓN DE INDICADORES

19. Meta 1. Enseñanza primaria y secundaria

Para 2030, velar porque todas las niñas y todos los niños terminen los ciclos de la enseñanza primaria y secundaria, que ha de ser gratuita, equitativa y de calidad y producir resultados escolares pertinentes y eficaces.

19.1. Indicador Mundial 1.1

Porcentaje de niños y de jóvenes: a) en los cursos segundo/tercero; b) al final de la enseñanza primaria; y c) al final del primer ciclo de la enseñanza secundaria, que alcanzan al menos un nivel mínimo de dominio de: i) la lectura y ii) las matemáticas, por sexos.

Observaciones:

- Actualmente no existen estándares comunes validos por la comunidad internacional para evaluar dominio mínimo de lectura y matemáticas.
- Medido a través de las evaluaciones de aprendizaje
- Comparación limitada entre países, la tabla 1 muestra los niveles adoptados por la UNESCO para comparar las diferentes pruebas de aprendizaje. Los cuadros naranjas representan los niveles mínimos.
- Se recomienda comparar entre los países que aplican la misma prueba
- Se limitará a evaluar las competencias de los niños y jóvenes que asisten a la escuela, se podría hacer pruebas a domicilio, pero puede ser costosa y difícil de administrar

Propósito:

El indicador es una medida directa de los resultados de aprendizaje alcanzados en las áreas evaluados en las etapas pertinentes de la educación.

Metodología de cálculo

El indicador se calcula como el porcentaje de niños o jóvenes en la fase pertinente de educación que alcanza o supera el nivel mínimo de competencia en una materia determinada.

$$\text{Rendimiento por encima del nivel mínimo, } PL_{tn, \text{mínimo alcanzado}} = p$$

Donde,

- p** Porcentaje de estudiantes en una evaluación de aprendizaje.
- n** Etapa de educación, a, b y c.
- s** Lectura o matemáticas
- t-i** El año de análisis t , donde i este entre 0 y 5.

Smin Desempeño mínimo en las pruebas

Notas.

- El estándar mínimo será definido por la comunidad educativa global teniendo en cuenta las diferencias regionales.
- Los tres puntos evaluados tendrán sus mínimos independientes, sólo se clasificarán entre los que no alcanzan el mínimo de competencia y los que sí.

Desagregación

Por edad o grupo de edad de los estudiantes, sexo, ubicación, situación socioeconómica, estatus migratorio y origen étnico. La situación de la discapacidad no está actualmente disponible en la mayoría de las evaluaciones de aprendizaje nacionales e internacionales.

Tabla 17. Niveles de competencia correspondientes a cada prueba internacional y regional de aprovechamiento estudiantil

Hamonized proficiency levels	Original name of levels by assessment																		
	PISA		TIMSS		PIRLS	PASEC				TERCE		SACMEQ							
	Grade and subject		Grade and subject		Grade and subject		Grade and subject		Grade and subject		Grade and subject								
	8	Math	4	Math	8	Math	4	Reading	2	2	6	6	3	3	6	6	6	6	
Level 8 (L8)	Level 6																	Level 8	Level 8
Level 7 (L7)	Level 5	Level 6																Level 7	Level 7
Level 6 (L6)	Level 4	Level 5																Level 6	Level 6
Level 5 (L5)	Level 3	Level 4	Advanced International Benchmark High	Advanced International Benchmark High	Advanced International Benchmark High	Level 4				Level 4								Level 5	Level 5
Level 4 (L4)	Level 2	Level 3	International Benchmark Intermediate	International Benchmark Intermediate	International Benchmark Intermediate	Level 3	Level 3	Level 3	Level 3	Level 3	Level 3	Level 3	Level 3	Level 3	Level 3	Level 3	Level 3	Level 4	Level 4
Level 3 (L3)	Level 1a	Level 2	International Benchmark Low	International Benchmark Low	International Benchmark Low	Level 2	Level 2	Level 2	Level 2	Level 2	Level 2	Level 2	Level 2	Level 2	Level 2	Level 2	Level 2	Level 3	Level 3
Level 2 (L2)	Level 1b	Level 1	International Benchmark Below Low	International Benchmark Below Low	International Benchmark Below Low	Level 1	Level 1	Level 1	Level 1	Level 1	Level 1	Level 1	Level 1	Level 1	Level 1	Level 1	Level 1	Level 2	Level 2
Level 1 (L1)	Below Level 1b	Below Level 1	International Benchmark Below Low	International Benchmark Below Low	International Benchmark Below Low	Level 1	Level 1	Level 1	Level 1	Level 1	Level 1	Level 1	Level 1	Level 1	Level 1	Level 1	Level 1	Level 1	Level 1

■ : achieve at least a minimum proficiency level

■ : not applicable

Fuentes:

UNESCO propone utilizar los resultados de las evaluaciones de aprendizaje - Pruebas TERCE.

Fuente	¿Qué es?	Disponibilidad	Formato	Observaciones
U - TERCE (Tercer Estudio Regional Comparativo y Explicativo)	El objetivo principal es obtener datos comparativos sobre el desempeño escolar de los estudiantes de grados tercero y sexto de América Latina y el Caribe en las áreas de Ciencias, Matemáticas y Lectura.	Bases de datos publicadas en la página web de la UNESCO. http://www.unesco.org/new/es/santiago/terce/databases/	SPSS, STATA, TXT	Se cuenta con la información del año 2013.
De: UNESCO		Bases de datos en el FTP del ICES. ftp://ftp.ices.gov.co/SABER11/SB11-BASES%20DE%20DATOS/		Para ingresar al FTP se necesita: Usuario: ftpdbicfes Contraseña: dataicfes%

19.2. Indicador Temático 1.2

Gestión de una evaluación del aprendizaje representativa nacionalmente: i) durante la enseñanza primaria, ii) al concluir la enseñanza primaria y iii) al concluir el primer ciclo de la enseñanza secundaria

Observaciones

- Evaluación nacional o transnacional de los resultados de aprendizaje de los últimos años en lectura, escritura y matemáticas en las etapas relevantes de la educación.
- La capacidad de los países para evaluar el aprendizaje mediante evaluaciones a gran escala es clave para supervisar la calidad y equidad de la educación.
- Si cumple con el indicador permitirá al país revisar y adaptarse políticas nacionales de educación y aprendizaje para garantizar que todos los niños y jóvenes la gente tiene la oportunidad de adquirir habilidades básicas en cada nivel de educación y en cada área temática.
- El indicador no mide habilidades sino existencia de evaluaciones de aprendizaje.

Metodología de calculo

El indicador se expresa simplemente con un SI o NO, para cada área temática y para cada etapa de la educación.

$LA_{n,s}^t=1$, si existe una evaluación de aprendizaje nacional o transnacional en cualquier año entre t-5 y t

Donde,

- LA Prueba de evaluación de aprendizaje
n Primaria (1) o secundaria inferior (2)
s Área temática
t El año de análisis t , donde i este entre 0 y 5.

Desagregación

Por nivel de educación y área temática

Fuentes

- Información sobre la aplicación de las evaluaciones de aprendizaje en cada materia y en cada etapa de la Educación.
- Datos sobre la administración de una evaluación a gran escala de una muestra representativa las oficinas nacionales de evaluación del aprendizaje, los ministerios de educación u otras evaluaciones de aprendizaje, incluidas las organizaciones regionales o internacionales que evaluaciones

Fuente	¿Qué es?	Disponibilidad	Formato
Ministerio de educación Nacional - Icfes	Lineamientos para las aplicaciones muestral y censal	http://www2.icfes.gov.co/instituciones-educativas-y-secretarias/pruebas-saber-3-5-y-9/guias-de-orientacion , Marco de factores asociados saber 3, 5 y 9 2016	Web

19.3. Indicador Temático 1.3

Tasas brutas de escolarización hasta el último curso (primaria, primer ciclo de secundaria)

Observaciones

- Número total de nuevos participantes en el último curso de primaria, independientemente de la edad. Expresado como porcentaje de la población prevista en edad de último curso de primaria.
- La edad de entrada está prevista para los estudiantes que ingresaron al sistema de educación en la edad oficial, estudian tiempo completo y no han progresado sin perder años.
- Es una medida aproximada de terminación de primaria. Refleja cómo el impacto de las políticas en el acceso y la progresión a través de los primeros grados de cada nivel de educación repercute en la calificación final de ese nivel.
- Indica la capacidad del sistema educativo para atender la edad de entrada prevista para el último grado del nivel dado de educación.
- Una proporción alta indica un alto grado de terminación de la educación primaria o secundaria inferior.
- Se trata de una medida bruta que puede superar el 100% si hay un gran número de alumnos que entraron a la escuela temprano o tarde y / o que han repetido grados anteriores
- En comparación con la tasa de terminación, la tasa bruta de consumo hasta el último grado no indica cuántos los niños completan el último grado, sólo cuántos niños entran en ese grado

Metodología de cálculo

El indicador se expresa como un porcentaje de la población de la edad de entrada prevista hasta el último grado de ese nivel de educación.

$$GIRLG_n \frac{NE_{l,n}}{P_{n,a}}$$

Donde,

$GIRLG_n$ Tasa de bruta

l Último grado primaria

n Primaria (1) o secundaria inferior (2)

$NE_{l,n}$ Nuevo entrantes al último grado l de nivel de educación n

$P_{n,a}$ Población de la edad a de ingreso prevista hasta el último grado de nivel n de educación

a Edad prevista para estudiar el último grado de primaria

Nota.

Si los datos sobre nuevos participantes no se recogen directamente, pueden calcularse restando el número de alumnos que repiten el último grado de la matrícula total en el último grado.

Desagregación

Por sexo y nivel educativo

Fuentes

- Datos administrativos sobre inscripción y repeticiones o nuevos participantes por grado. (SIMAT)
- Población de censos
- Datos del ministerio de educación sobre la estructura del sistema educativo

Fuente	¿Qué es?	Disponibilidad	Formato	Observaciones
SIMAT, Sistema integrado de matriculados De: Ministerio de Educación Nacional de Colombia	Es una herramienta que permite organizar y controlar el proceso de matrícula en todas sus etapas, así como tener una fuente de información confiable y disponible para la toma de decisiones.	Restringido	No aplica	
Proyecciones de población De: Departamento Administrativo Nacional de Estadística - DANE	Estas proyecciones fueron realizadas tomando como base los resultados ajustados de población del Censo 2005 y la conciliación censal 1985 - 2005; así como los análisis sobre el comportamiento de las variables determinantes de la evolución demográfica, las hipótesis y algunos comentarios sobre sus resultados	Archivo de proyecciones disponible en la página web del DANE. http://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion	Excel	Es una tabla dinámica que permite seleccionar municipios, años, sexo y genera la población estimada.

19.4. Indicador Temático 1.4

Tasas de terminación de estudios (primaria, primer ciclo de secundaria, segundo ciclo de secundaria)

Observaciones

- Porcentaje de una cohorte de niños o jóvenes de 3 a 5 años por encima de la edad prevista para el último grado de cada nivel de educación que han completado ese grado
- La edad de entrada está prevista para los estudiantes que ingresaron al sistema de educación en la edad oficial, estudian tiempo completo y no han progresado sin perder años. La tasa de finalización indica cuántas personas en un grupo de edad dado han completado el nivel de educación.
- Una tasa de terminación cercana al 100% indica que la mayoría o todos los niños y adolescentes han completado
- Una tasa de terminación baja indica una entrada baja o retrasada en un determinado nivel de educación, una alta deserción escolar, alta repetición, finalización tardía, o una combinación de estos factores.
- Para identificar las causas de las bajas tasas de terminación, es necesario examinar otros indicadores, por ejemplo, la tasa de abandono escolar, la tasa bruta de ingesta al último grado y el porcentaje de niños mayores de edad. Cuando se desglosan por sexo, ubicación y otras características, este indicador puede identificar grupos de población.

Metodología de cálculo

El número de personas en el grupo de edad relevante que han completado el último grado del nivel dado de la educación se expresa como un porcentaje de la población total del mismo grupo de edad

$$CR_n = \frac{EAP_{n,AG(a+3t5)}}{PAG_{(a+3t5)}}$$

Donde,

CR_n	Tasa de terminación del nivel de educación
$EAP_{n,AG(a+3t5)}$	Población de 3 a 5 años por encima de la edad de entrada oficial a en el último grado de nivel n de educación que completó el nivel n
$PAG_{(a+3t5)}$	Población de 3 a 5 años por encima de la entrada oficial a en el último grado de nivel n de educación.

Desagregación

Por edad o grupo de edad de los estudiantes, sexo, ubicación, estado socioeconómico, nivel de educación y otros como disponibles en encuestas o datos censales

Fuentes

- Población en el grupo de edad pertinente por el nivel más alto de educación o grado completado; datos sobre la estructura (edad de entrada y duración) de cada nivel de educación

- Población de censos
- Encuestas de estimaciones de población por edad
- Datos del ministerio de educación sobre la estructura del sistema educativo
- Las encuestas sobre la fuerza de trabajo pueden servir como fuente de datos para la terminación secundaria inferior y superior si recogen información para los grupos de edad de interés
- Las encuestas internacionales de muestreo, como las encuestas demográficas y de salud DHS

Fuente	¿Qué es?	Disponibilidad	Formato	Observaciones
SIMAT, Sistema integrado de matriculados De: Ministerio de Educación Nacional de Colombia (MEN)	Es una herramienta que permite organizar y controlar el proceso de matrícula en todas sus etapas, así como tener una fuente de información confiable y disponible para la toma de decisiones.	Restringido	No aplica	
Proyecciones de población De: Departamento Administrativo Nacional de Estadística – (DANE)	Estas proyecciones fueron realizadas tomando como base los resultados ajustados de población del Censo 2005 y la conciliación censal 1985 - 2005; así como los análisis sobre el comportamiento de las variables determinantes de la evolución demográfica, las hipótesis y algunos comentarios sobre sus resultados	Archivo de proyecciones disponible en la página web del DANE. http://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion	Excel	Es una tabla dinámica que permite seleccionar municipios, años, sexo y genera la población estimada.
DHS, Encuesta de demografía y Salud. De: Profamilia	La Encuesta Nacional de Demografía y Salud es un documento que Profamilia publica cada cinco años desde 1990 en el que se recoge información sobre la dimensión, estructura, evolución, dinámicas y características generales de la población colombiana en el ámbito de salud sexual y reproductiva y otros temas relativos a la salud.	Bases de datos disponibles en la página web de la Encuesta http://dhsprogram.com	SPSS	Se debe crear un usuario y describir el proyecto y el objetivo de las bases de datos. Las bases ya están descargadas.

19.5. Indicador Temático 1.5

Tasas de niños y adolescentes fuera de la escuela (primaria, primer ciclo de secundaria, segundo ciclo de secundaria)

Observaciones

- Los niños y los jóvenes en la franja de edad oficial para el nivel de educación dado que no están matriculados en los niveles de educación primaria, secundaria o superior
- Los niños y jóvenes que se matriculan en la educación preescolar se consideran fuera de la escuela.
- Pretende identificar el tamaño de la población en el rango de edad oficial para el nivel de educación dado que no están matriculados en la escuela con el fin de que puedan ser mejor orientados y políticas adecuadas pueden ser puestas en marcha para asegurar que tengan acceso a la educación.
- Cuanto mayor es el número de niños y adolescentes fuera de la escuela, mayor es la necesidad de enfocarse en mejorar el acceso a la educación.
- Cuando se desglosan por sexo, ubicación y otras características, este indicador puede identificar grupos excluidos de la población

Metodología de cálculo

El número de estudiantes de la edad oficial para el nivel dado de educación matriculado en los niveles de educación primaria, secundaria o superior es restado de la población total de la misma edad

$$OSR_n = \frac{SAP_n - \sum_{i=1}^8 E_{i,AGn}}{SAP_N}$$

Donde,

OSR_n Tasa fuera de la escuela para niños y jóvenes de la edad oficial para el nivel n de educación

SAP_n Población de la edad oficial para el nivel n de educación

E_{i,AGn} Inscripción en el ISCED i de los niños y jóvenes de la edad oficial para el nivel n de educación

Desagregación

Por edad o grupo de edad de los estudiantes, sexo, ubicación, estado socioeconómico, nivel de educación y otros como disponibles en encuestas o datos censales

Fuentes

- Población en el grupo de edad pertinente por el nivel más alto de educación o grado completado; datos sobre la estructura (edad de entrada y duración) de cada nivel de educación
- Población de censos
- Datos del ministerio de educación sobre la estructura del sistema educativo

Fuente	¿Qué es?	Disponibilidad	Formato	Observaciones
SIMAT, Sistema integrado de matriculados De: Ministerio de Educación Nacional de Colombia	Es una herramienta que permite organizar y controlar el proceso de matrícula en todas sus etapas, así como tener una fuente de información confiable y disponible para la toma de decisiones.	Restringido	No aplica	
Proyecciones de población De: Departamento Administrativo Nacional de Estadística - DANE	Estas proyecciones fueron realizadas tomando como base los resultados ajustados de población del Censo 2005 y la conciliación censal 1985 - 2005; así como los análisis sobre el comportamiento de las variables determinantes de la evolución demográfica, las hipótesis y algunos comentarios sobre sus resultados	Archivo de proyecciones disponible en la página web del DANE. http://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion	Excel	Es una tabla dinámica que permite seleccionar municipios, años, sexo y genera la población estimada.
Datos abiertos, gobierno digital Colombia. De: MEN	Los datos abiertos hacen parte de la información pública y representan un gran potencial para el desarrollo de procesos de Gobierno Abierto (Open Government)	Página de datos abiertos https://www.datos.gov.co/Educacion/ESTADISTICAS-EN-EDUCACION-BASICA-POR-ETC/sras-4t5p/data	En línea, calculado MEN	El ministerio de Educación dispone de una página web con tasas de cobertura calculadas desde 2011 hasta 2015 por departamento.

19.6. Indicador Temático 1.6

Porcentajes de niños de edad superior al curso que estudian (primaria, primer ciclo de secundaria)

Observaciones

- Porcentaje de alumnos de cada nivel educativo (enseñanza primaria y secundaria inferior) que estén por lo menos 2 años por encima de la edad prevista para su grado
- El indicador mide el progreso hacia asegurar que todas las niñas y niños completen un ciclo completo de educación primaria y secundaria y lograr al menos niveles mínimos de competencia en lectura y matemáticas en cada nivel
- Los niños pueden estar sobre-edad para una calificación porque comenzaron tarde de la escuela y / o han repetido uno o más grados anteriores
- Un valor bajo de este indicador demostrará que la mayoría de los estudiantes comienzan la escuela a tiempo y progresan con niveles mínimos de repetición de grado
- La progresión de la sobre-edad y la repetición significativa deben ser desalentados ya que ambos están asociados con niveles más bajos de logro de aprendizaje de los estudiantes.

Metodología de cálculo

La suma de matrículas en todos los grados en el nivel de educación que son 2 o más años mayor a la edad prevista para el grado dado se expresa como un porcentaje del total de la matrícula en el nivel dado de educación.

$$POAG_n = \frac{\sum_{g=1}^{dn} dn E_{n,g,AG,2+}}{E_n}$$

Donde,

$POAG_n$	Porcentaje de niños mayores de edad para el grado en el nivel n de educación
$E_{n,g,AG,2+}$	Matriculados en el grado g de nivel n de educación que tengan por lo menos dos años más de la edad prevista para ese grado
E_n	Matrícula total en el nivel n de educación
d_n	Duración (en años) del nivel n de educación
n	Primaria (1) o secundaria inferior (2)

Desagregación

Por edad o grupo de edad de los estudiantes, sexo, ubicación, estado socioeconómico, nivel de educación y otros como disponibles en encuestas o datos censales

Fuentes

- Población en el grupo de edad pertinente por el nivel más alto de educación o grado completado;
- datos sobre la estructura (edad de entrada y duración) de cada nivel de educación
- Población de censos

- Datos del ministerio de educación sobre la estructura del sistema educativo

Fuente	¿Qué es?	Disponibilidad	Formato	Observaciones
SIMAT, Sistema integrado de matriculados De: MEN	Es una herramienta que permite organizar y controlar el proceso de matrícula en todas sus etapas, así como tener una fuente de información confiable y disponible para la toma de decisiones.	Restringido	No aplica	
Proyecciones de población De: DANE	Estas proyecciones fueron realizadas tomando como base los resultados ajustados de población del Censo 2005 y la conciliación censal 1985 - 2005; así como los análisis sobre el comportamiento de las variables determinantes de la evolución demográfica, las hipótesis y algunos comentarios sobre sus resultados	Archivo de proyecciones disponible en la página web del DANE. http://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion	Excel	Es una tabla dinámica que permite seleccionar municipios, años, sexo y genera la población estimada.

19.7. Indicador Temático 1.7

Número de años de: enseñanza primaria y secundaria i) gratuita y ii) obligatoria que garantizan los marcos jurídicos

Observaciones

- Número de años de educación primaria y secundaria a los que tienen derecho los niños y los jóvenes, que están libres de matrícula o son obligatorios o ambos.
- Pretende medir el compromiso del gobierno de garantizar el derecho a la educación de los niños y jóvenes
- La existencia de una legislación nacional que garantice el derecho a la educación en determinadas edades y / o grados demuestra el compromiso del gobierno de garantizar que los niños y los jóvenes regularmente
- Cuanto mayor sea el número de años garantizados, más probable es que los niños y jóvenes las personas permanecerán más tiempo en la escuela y tendrán la oportunidad de adquirir las competencias en cada nivel de educación.
- La existencia de la legislación no asegura la implementación de los países

Metodología de cálculo

Registra el número de grados de educación primaria y secundaria garantizados

YF₁₂₃ Número de años de educación primaria y secundaria gratuita (niveles ISCED 1, 2 y 3)

YC₁₂₃ (Niveles ISCED 1, 2 y 3)

Desagregación

Nivel de educación

Fuentes

- Legislación nacional y normas de educación formal sobre el acceso a la escolaridad y, en particular, el derecho legal u obligación de asistir a la escuela; y datos administrativos de los ministerios de educación sobre la estructura del sistema educativo.

Fuente	¿Qué es?	Disponibilidad	Formato
Decreto 4807 del 20 de diciembre de 2011	Por el cual se define, toda la primaria el bachillerato hasta grado 11 gratuita	NA	Web
Constitución nacional artículo 67	Define toda la primaria el bachillerato hasta grado 11 obligatoria	NA	Web

20. Meta 2. Primera infancia

En torno a 2030, conseguir que todas las niñas y los niños tengan acceso a un desarrollo de su primera infancia, cuidado y educación preescolar de calidad con el fin de que estén listos para recibir la educación primaria.

20.1. Indicador Mundial 2.1

Proporción de niños menores de cinco años que se encuentran en camino de desarrollo que les corresponde en lo relativo a salud, aprendizaje y bienestar psicosocial, por sexos

Observaciones:

- Aun no existe una definición globalmente aceptada de "on track" (en camino de..).
- Definir y medir en camino de desarrollo debe ser una meta para la próxima fase de desarrollo de este indicador.
- Actualmente la MICS mide el camino en desarrollo como el niño que:
 - *Aprendizaje*, identifica al menos 10 letras del alfabeto, lee 4 palabras simples, reconoce y lee los números de 1 a 10. Participa en cualquier tipo de aprendizaje, guardería, jardín, casa comunitaria.
 - *Físicamente*, recoge fácilmente objetos pequeños y son suficientes para jugar
 - *Socio-emocional*, capaces de hacer tareas sencillas de manera independientes, llevarse bien con otros niños, no patear, morder o golpear otros niños y adultos.
- Un valor alto indica que un gran número de niños pequeños están bien preparados para la escuela en las áreas de salud, aprendizaje y bienestar psicosocial.
- Se necesitará más trabajo de desarrollo metodológico para asegurar que la medida propuesta refleje una definición comúnmente acordada

Propósito:

Es una medida amplia del desarrollo de los niños y su preparación para ingresar al colegio.

Metodología de cálculo

Registra el porcentaje de niños entre 36-59 meses que demuestran niveles apropiados de desarrollo según la edad en las áreas medidas de acuerdo a los ítems usados en las MICS.

$$PCDT_{3t4} = \frac{CTD_{3t4}}{P_{3t4}}$$

Donde,

$PCDT_{3T4}$ Porcentaje de niños de 3 a 4 años de edad (36-59 meses) que son mentalmente desarrollados respecto a salud, aprendizaje y buen comportamiento psicosocial.

CDT_{3T4} Niños de 3 a 4 años de edad (36-59 meses) que son mentalmente desarrollados respecto a salud, aprendizaje y buen comportamiento psicosocial.

E_n Población de niños de 3 a 4 años de edad (36-59 meses).

Desagregación

Por edad, sexo, ubicación, ingresos, antecedentes familiares y participación en programas de educación de la primera infancia

Fuentes:

- Encuesta de hogares o encuesta directa a niños.
- Se necesita una fuente que diga el número de niños entre 36 y 59 meses que muestran niveles de desarrollo adecuados para su edad en las áreas que se están midiendo y el número total de niños en el mismo grupo
- Índice de desarrollo de la primera infancia
- Encuesta MICS
- PRIDI

Fuente	¿Qué es?	Disponibilidad	Formato	Observaciones
GEIH, Gran Encuesta Integrada de Hogares. De: DANE	La encuesta se ha especializado en la medición de la estructura del mercado laboral y los ingresos de los hogares. Esta tiene una muestra total anual de 240.000 hogares aproximadamente, lo que hace que sea la de mayor cobertura a nivel nacional	https://formularios.dane.gov.co/Anda_4_1/index.php/catalog/458/get_mic_rodata No incluye lo temas de primera infancia necesarios.	Txt	Sin embargo para futuros análisis, en el link relacionado reposan los microdatos desde 2001 hasta 2017.
MICS, Indicator Cluster Survey. De: UNICEF	Durante más de dos décadas, se han llevado a cabo cerca de 300 Encuestas de Indicadores Múltiples por Conglomerados en más de 100 países, generando datos sobre indicadores clave sobre el bienestar de los niños y las mujeres, y ayudando a diseñar políticas para mejorar sus vidas.	http://mics.unicef.org/surveys Colombia no participó en estas encuestas.	Online	Sin embargo para futuros análisis, en el link relacionado se presentan resultados

PRIDI, Programa Regional de Indicadores de Desarrollo Infantil. De: Banco Interamericano de desarrollo.	Es un programa regional que recopiló y analizó datos e indicadores comparables sobre el desarrollo de la primera infancia (ECD) en América Latina a fin de obtener una visión más profunda de la situación de los niños pequeños y sus familias	http://www.iea.nl/data Colombia no participó en estos indicadores.	Online	Sin embargo para futuros análisis, en el link relacionado se presentan resultados
Proyecciones de población	Estas proyecciones fueron realizadas tomando como base los resultados ajustados de población del Censo 2005 y la conciliación censal 1985 - 2005; así como los análisis sobre el comportamiento de las variables determinantes de la evolución demográfica, las hipótesis y algunos comentarios sobre sus resultados	Archivo de proyecciones disponible en la página web del DANE. http://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion	Excel	Es una tabla dinámica que permite seleccionar municipios, años, sexo y genera la población estimada.

20.2. Indicador Mundial 2.2

Tasa de participación en un aprendizaje organizado (un año antes de la edad oficial de ingreso en la enseñanza primaria), por sexos

Observaciones:

- Porcentaje de niños que participan en uno o más programas que combinan atención y educación
- Participación en la educación de la primera infancia
- La edad varía por cada país según la edad oficial para ingresar al sistema educativo
- Un programa de aprendizaje organizado consiste en un conjunto coherente o secuencia de actividades educativas diseñadas con la intención de lograr resultados de aprendizaje predeterminados o la realización de un conjunto específico de tareas educativas
- Un alto valor del indicador muestra un alto grado de participación en el aprendizaje organizado inmediatamente antes de la edad de ingreso oficial a la educación primaria
- El indicador mide el porcentaje de niños que están expuestos al aprendizaje organizado, pero no la intensidad del programa, que limita la capacidad de sacar conclusiones sobre la medida en que se alcanza este objetivo.
- Falta cantidad de tiempo que los niños pasan en los programas de aprendizaje

Propósito:

El indicador mide la exposición de los niños a actividades de aprendizaje organizadas en el año el comienzo de la escuela primaria

Metodología de cálculo

El número de niños con edad relevante que participan en un aprendizaje organizado, es expresado como el porcentaje del total de la población en el mismo rango de edad.

$$PROL_{0t1,AG(a-1)} = \frac{E_{0t1,AG(a-1)}}{SAP_{AG(a-1)}}$$

Donde,

$PROL_{0t1,AG(a-1)}$	Tasa de participación en aprendizaje organizado un año antes de la edad oficial de ingreso en la enseñanza primaria.
$E_{0t1,AG(a-1)}$	Inscripción en la primera infancia o educación primaria (ISCED niveles 0 y 1 con un año de edad inferior a la edad oficial de ingreso para la educación primaria.
$SAP_{AG(a-1)}$	Edad escolar de la población con un año de edad inferior a la edad oficial de ingreso a la educación primaria.

Desagregación

Por edad y sexo de fuentes administrativas, y por edad, sexo, ubicación e ingresos de las encuestas de hogares.

Fuentes:

- Se necesita una fuente que diga el número de niños que participan en actividades organizadas de aprendizaje por año de edad; población estimaciones por año de edad (si se utilizan datos administrativos); y datos sobre la entrada oficial edad a la educación primaria.
- Datos administrativos de escuelas y otros centros de aprendizaje organizado o de encuestas de hogares sobre la inscripción por año de edad en programas de aprendizaje temprano
- Censos de población y encuestas para estimaciones de población por año de edad (si se utilizan datos administrativos sobre la matrícula)

Fuente	¿Qué es?	Disponibilidad	Formato	Observaciones
Investigación de Educación Formal 2007 - 2016. De: DANE	Censo de establecimientos educativos que ofrecen educación básica formal en los niveles educativos de preescolar, básica primaria, básica secundaria y media. Se incluye también ciclos lectivos integrados (educación para adultos y jóvenes en extra edad) y otros modelos educativos	Acceso a microdatos del DANE. http://formularios.dane.gov.co/Anda_4_1/index.php/catalog/503/get_microdata	TXT	Se debe crear usuario, explicando el uso de los datos.

Proyecciones población	de	Estas proyecciones fueron realizadas tomando como base los resultados ajustados de población del Censo 2005 y la conciliación censal 1985 - 2005; así como los análisis sobre el comportamiento de las variables determinantes de la evolución demográfica, las hipótesis y algunos comentarios sobre sus resultados	Archivo de proyecciones disponible en la página web del DANE. http://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion	Excel	Es una tabla dinámica que permite seleccionar municipios, años, sexo y genera la población estimada.
------------------------	----	--	--	-------	--

20.3. Indicador Temático 2.3

Porcentaje de niños menores de cinco años que gozan de entornos de aprendizaje familiares positivos y estimulantes

Observaciones:

- Porcentaje de niños de entre 36 y 59 meses que viven en hogares donde su madre, padre u otros miembros adultos del hogar se involucran con ellos en los siguientes tipos de actividades: leer o mirar libros ilustrados; contando historias; cantando canciones; llevar a los niños fuera del hogar; jugando y nombrando, contando y / o dibujando.

Propósito:

- Dentro de la casa, los cuidadores tienen la tarea de establecer un ambiente seguro, estimulante y nutritivo y proporcionar dirección y orientación en la vida cotidiana. Las interacciones con cuidadores responsables que son sensibles y responden a las habilidades emergentes de los niños son fundamentales para el desarrollo social, emocional y cognitivo. Este tipo de cuidado positivo puede ayudar a los niños a sentirse valoradas y aceptadas, promueven reacciones saludables, proporcionan un modelo de relaciones, y contribuir al éxito posterior académico y del empleo.
- Este indicador ofrece una amplia medida de las formas en que los adultos en el hogar interactúan con los niños en formas significativas y estimulantes para promover el aprendizaje y la preparación escolar.

Metodología de cálculo

Refleja el porcentaje de niños entre los 36 y 59 meses que participan en actividades en las áreas medidas.

$$PCPSH_{3t4} = \frac{CPSH_{3t4}}{P_{3t4}}$$

Donde,

$PCPSH_{3t4}$ Porcentaje de niños entre 3 y 4 años que experimentan entornos de aprendizaje familiares positivos y estimulantes.

$CPSH_{3t4}$ Niños de 3 a 4 años que experimentan entornos de aprendizaje familiares positivos y estimulantes.

P_{3t4} Población de niños entre 3 y 4 años de edad.

Desagregación

Por edad, sexo, ubicación e ingresos.

Fuentes:

- Encuesta de hogares
- Una fuente que relacione el número de niños de entre 36 y 59 meses que participan en actividades en las medidas y el número total de niños en el mismo grupo de edad.
- Medidas de ambientes de aprendizaje hogareño positivos y estimulantes para niños pequeños que se han utilizado en varios países están disponibles en las encuestas y evaluaciones,
- Encuestas de Indicadores de Indicadores Múltiples, PRIDI en América Latina

Fuente	¿Qué es?	Disponibilidad	Formato	Observaciones
Encuesta de calidad de vida. 2015-2016. De: DANE	Las encuestas de calidad de vida surgen como respuesta a la necesidad de caracterizar la población en los diferentes aspectos involucrados en el bienestar de los hogares. Desde hace pocos años se ha incluido un módulo: Atención integral de los niños y niñas menores de 5 años.	Acceso a microdatos del DANE. http://formularios.dane.gov.co/Anda_4_1/index.php/catalog/456/get_microdata	TXT	Se debe crear usuario, explicando el uso de los datos.
Proyecciones de población	Estas proyecciones fueron realizadas tomando como base los resultados ajustados de población del Censo 2005 y la conciliación censal 1985 - 2005; así como los análisis sobre el comportamiento de las variables determinantes de la evolución demográfica, las hipótesis y algunos comentarios sobre sus resultados	Archivo de proyecciones disponible en la página web del DANE. http://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion	Excel	Es una tabla dinámica que permite seleccionar municipios, años, sexo y genera la población estimada.

20.4. Indicador temático 2.4

Tasa bruta de matriculación en enseñanza preescolar

Observaciones:

- Total, de la matrícula en la educación preescolar independientemente de la edad expresada como porcentaje de la población de la edad oficial para la educación preescolar.
- Un alto valor indica generalmente un alto grado de participación, ya sea que los alumnos pertenezcan al grupo de edad oficial o no. Un valor aproximado o superior al 100% indica que un país es capaz de acomodar a toda su población en edad pre-primaria, pero no indica la proporción ya inscrita

Propósito:

- Muestra el nivel general de participación en la educación preescolar. Indica la capacidad del sistema educativo para matricular a los estudiantes de un grupo de edad en particular.

Desagregación

Por edad, sexo, ubicación e ingresos.

Fuentes:

- Encuesta de hogares
- Inscripción en la educación preescolar; estimaciones de la población por año de edad (si se utilizan datos administrativos) y datos sobre la estructura (edad de entrada y duración) de la educación preescolar educación.
- Datos administrativos de las escuelas o datos de encuestas de hogares sobre inscripción; censos de población y encuestas para estimaciones de población por año de edad (si se utilizan datos administrativos sobre inscripción); datos administrativos de los ministerios de educación sobre la estructura de la educación temprana de la niñez.

Fuente	¿Qué es?	Disponibilidad	Formato	Observaciones
Investigación de Educación Formal 2007 - 2016. De: DANE	Censo de establecimientos educativos que ofrecen educación básica formal en los niveles educativos de preescolar, básica primaria, básica secundaria y media. Se incluye también ciclos lectivos integrados (educación para adultos y jóvenes en extra edad) y otros modelos educativos	Acceso a microdatos del DANE. http://formularios.dane.gov.co/Anda_4_1/index.php/catalog/503/get_microdata	TXT	Se debe crear usuario, explicando el uso de los datos.

Proyecciones de población	de	Estas proyecciones fueron realizadas tomando como base los resultados ajustados de población del Censo 2005 y la conciliación censal 1985 - 2005; así como los análisis sobre el comportamiento de las variables determinantes de la evolución demográfica, las hipótesis y algunos comentarios sobre sus resultados	Archivo de proyecciones disponible en la página web del DANE. http://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion	Excel	Es una tabla dinámica que permite seleccionar municipios, años, sexo y genera la población estimada.
---------------------------	----	--	--	-------	--

20.5. Indicador temático 2.5

Número de años de: enseñanza preescolar i) gratuita y ii) obligatoria que garantizan los marcos jurídicos

Observaciones:

- Número de años de educación preescolar a los que tienen derecho los niños legalmente libres de tasas de matrícula u obligatorias o ambas.
- El número de años de educación preescolar a los que tienen derecho los niños debería ser idealmente el número de grados de educación preescolar que se espera que los niños tengan completado antes de entrar en la educación primaria.
- Cuanto mayor sea el número de años de educación preescolar más probabilidades de que los niños estén bien preparados para ingresar a la educación primaria tiempo apropiado.

Propósito:

Medir el compromiso del gobierno de garantizar el derecho a la educación de los niños y gente joven.

Metodología de cálculo

Consiste en grabar el número de grados de la educación preescolar que están garantizados. Si se usa el parámetro edades en lugar que grados, sustraer la edad más baja de la edad oficial para entrar a la escuela primaria. Si el resultado es 0 o negativo, no hay años de preescolar que estén garantizados.

YF₀₂ Número de años de educación preescolar gratuitos. (ISCED nivel 02).

YC₀₂ Número de años de educación preescolar gratuitos. (ISCED nivel 02).

Desagregación

NA.

Fuentes:

- Datos administrativos
- Fuente con el número de grados de educación preescolar que están (a) libres de tasas de matrícula y / o b) obligatorias según la legislación nacional. Si no se especifica el número de calificaciones, la edad en el que la educación es (a) gratuita y / o (b) obligatoria
- La legislación nacional y las normas de educación formal sobre el acceso a la en particular, el derecho legal u obligación de asistir a la escuela; datos administrativos de ministerios de educación sobre la estructura del sistema educativo.
- La existencia de una legislación nacional no garantiza que los países que los padres están efectivamente asegurando que sus hijos se beneficien de la disposición disponible.

Fuente	¿Qué es?	Disponibilidad	Formato
Decreto 4807 del 20 de diciembre de 2011	Por el cual se define, toda la primaria el bachillerato hasta grado 11 gratuita	NA	Web
Constitución nacional artículo 67	Define toda la primaria el bachillerato hasta grado 11 obligatoria	NA	Web

21. Meta 3. Enseñanza técnica, profesional, superior y de adultos

Para 2030, asegurar el acceso, en condiciones de igualdad para todos los hombres y las mujeres, a una formación técnica, profesional y superior de calidad, incluida la enseñanza universitaria

21.1. Indicador Mundial 3.1

Tasa de participación de los jóvenes y adultos en la enseñanza formal y no formal en los 12 meses anteriores, por sexos

Observaciones:

- Porcentaje de jóvenes y adultos en un rango de edad determinado en la educación formal o no formal o la formación en un período de tiempo determinado
- Idealmente, el indicador debería desagregarse por tipos de programas como la EFTP, la educación terciaria, la educación de adultos y otros tipos pertinentes y cubren programas formales y no formales.
- La educación no formal tiene lugar dentro y fuera de las instituciones educativas y atender a las personas de todas las edades.
- La educación no formal dependiendo de los contextos nacionales, puede cubrir programas educativos para impartir alfabetización de adultos, habilidades para la vida, habilidades laborales y cultura general.
- Un valor elevado indica que una gran parte de la población del grupo de edad pertinente participa en educación formal y no formal.

Propósito: Mostrar el nivel de participación de jóvenes y adultos en la educación y la formación de todo tipo

Metodología de cálculo

El número de gente en grupos por edades seleccionados que participan en educación formal o informal. Es expresado como el porcentaje de la población de la misma edad.

$$PR_{AGi} = \frac{E_{AGi}}{P_{AGi}}$$

Donde,

PR_{AGi} Tasa de participación de la población en el grupo de edad i en educación formal o informal y formación.

E_{AGi} Inscripción de la población en el grupo de edad i en educación formal o informal y formación.

i 15-24 años, 15 años y más, 25-64 años etc.

Desagregación
Por edad, sexo,

Fuentes:

- Encuesta de hogares o datos administrativos.
- Se necesita una fuente que diga el número de participantes por año de edad en la educación y la formación formales y no formales y las estimaciones de población por año de edad.
- Datos administrativos de escuelas y otros lugares de educación y formación o datos de encuestas de hogares sobre los participantes en la educación formal y no formal y la formación por año de edad
- Población censos y encuestas para las estimaciones de población por año de edad

21.2. Indicador Temática 3.2

Tasa bruta de matriculación en la enseñanza superior

Observaciones:

- Total, de la matrícula en educación terciaria, independientemente de la edad, expresada como porcentaje de la población en el grupo de 5 años de edad inmediatamente después de la educación secundaria superior.
- Un alto valor del indicador muestra un alto grado de participación en la educación terciaria todas las edades.
- La tasa bruta de matrícula es una medida amplia de la participación en la educación terciaria y no las diferencias en la duración de los programas entre países o entre distintos niveles de educación y campos de estudio

Propósito:

Mostrar el nivel general de participación en un determinado nivel de educación. Indica la capacidad del sistema educativo para matricular a los estudiantes de un grupo de edad determinado.

Metodología de cálculo

Indica el número de estudiantes matriculados en educación terciaria, expresado como el porcentaje del grupo de 5 años de edad inmediatamente después de la educación secundaria.

$$GER_{5t8} = \frac{E_{5t8}}{SAP_{5t8,a}}$$

Donde,

GER_{5t8} Tasa bruta de matriculación en educación terciaria (ISCED niveles 5, 6, 7 y 8).

E_{5t8} Matriculación en educación terciaria (ISCED niveles 5, 6, 7 y 8).

SAP_{5t8,a} Población con la edad oficial (**a**) para la educación terciaria (ISCED niveles 5, 6, 7 y 8).

Desagregación

Por sexo de fuentes administrativas, y por sexo, ubicación e ingresos de las encuestas de hogares.

Fuentes:

- Encuesta de hogares o datos administrativos.
- Se necesita una fuente que diga los matriculados en la educación terciaria; estimaciones de la población por año de edad y datos sobre la estructura (edad de entrada y duración) de la educación secundaria superior.
- Población censos y encuestas para las estimaciones de población por año de edad

Fuente	¿Qué es?	Disponibilidad	Formato	Observaciones
SNIES. Sistema Nacional de Información de la Educación Superior. 2014 - 2016. De: MEN	Es un sistema de información que ha sido creado para responder a las necesidades de información de la educación superior en Colombia.	Acceso a agregados por institución. http://www.mineducacion.gov.co/sistemasdeinformacion/1735/w3-article-212400.html	Excel	Acceso libre sin usuarios.
Proyecciones de población	Estas proyecciones fueron realizadas tomando como base los resultados ajustados de población del Censo 2005 y la conciliación censal 1985 - 2005; así como los análisis sobre el comportamiento de las variables determinantes de la evolución demográfica, las hipótesis y algunos comentarios sobre sus resultados	Archivo de proyecciones disponible en la página web del DANE. http://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion	Excel	Es una tabla dinámica que permite seleccionar municipios, años, sexo y genera la población estimada.

21.3. Indicador Temática 3.3

Tasa de participación en programas de enseñanza técnica y profesional (personas de 15 a 24 años de edad)

Observaciones:

- Porcentaje de jóvenes de 15 a 24 años que participan en la enseñanza técnica o profesional, ya sea en la educación formal, en el trabajo o en otros entornos, en una fecha determinada o durante un período determinado.
- Un valor alto indica que gran parte de la población de 15 a 24 años de edad está participando en la educación y la formación diseñadas específicamente para conducir a un trabajo.
- Las tasas de participación no capturan la intensidad o calidad de la provisión ni los resultados de la educación y la formación que se ofrecen.

Propósito: Mostrar el nivel de participación de los jóvenes en la educación y la formación técnica y profesional.

Metodología de cálculo

Expresado como el porcentaje de la población del mismo grupo de edad.

$$PR_{V,15T24} = \frac{E_{y,15t24}}{P_{15t24}}$$

Donde,

$PR_{V,15T24}$ Tasa de participación de jóvenes entre 15 a 24 años en educación técnica y profesional.

$E_{V,15t24}$ Matriculación en educación técnica y profesional de gente joven entre 15 y 24 años de edad.

P_{15t24} Población de 15 a 24 años de edad.

Desagregación

Por sexo de fuentes administrativas, y por sexo, ubicación e ingresos de las encuestas de hogares.

Fuentes:

- Encuesta de hogares o datos administrativos.
- Población censos y encuestas para las estimaciones de población por año de edad
- Una fuente que responda al número de participantes de 15 a 24 años de edad en la enseñanza y la formación técnica y profesional; estimaciones de población por año de edad.

Fuente

¿Qué es?

Disponibilidad

Formato

Observaciones

SNIES. Sistema Nacional de Información de la Educación Superior. 2014 - 2016. Es un sistema de información que ha sido creado para responder a las necesidades de información de la educación superior en Colombia. Acceso a agregados por institución. Excel Acceso libre sin usuarios.

<http://www.mineducacion.gov.co/sistemasdeinformacion/1735/w3-article-212400.html>

De: MEN

Proyecciones de población de Estas proyecciones fueron realizadas tomando como base los resultados ajustados de población del Censo 2005 y la conciliación censal 1985 - 2005; así como los análisis sobre el comportamiento de las variables determinantes de la evolución demográfica, las hipótesis y algunos comentarios sobre sus resultados Archivo de proyecciones disponible en la página web del DANE. Excel Es una tabla dinámica que permite seleccionar municipios, años, sexo y genera la población estimada.

<http://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion>

22. Meta 4. Competencias para el trabajo

Aumentar considerablemente el número de jóvenes y adultos que tienen las competencias necesarias, en particular técnicas y profesionales, para acceder al empleo, el trabajo decente y el emprendimiento.

22.1. Indicador Mundial 4.1

Porcentaje de jóvenes y adultos con competencias en tecnología de la información y las comunicaciones (TIC), por tipos de competencia.

Observaciones:

- Algunas de las actividades relacionadas con los computadores para medir habilidades sobre las TIC son:
 - *Uso de herramientas:*
 - Copiar y pegar para duplicar información.
 - Copiar y mover documentos o carpetas.
 - Fórmulas de aritmética básica en hojas de cálculo.
 - Presentaciones electrónicas con textos, imágenes, sonido y video.
 - Nuevos dispositivos como cámaras, impresoras, modem.
 - *Envío de correos electrónicos*, con archivos adjuntos.
 - *Encontrar, descargar, instalar y configurar software.*
 - *Uso de lenguaje especializado para programación.*
- Un computador se refiere a una máquina de escritorio, portable o a una tableta. No incluye el manejo por ejemplo de un televisor inteligente o de un teléfono móvil.

Propósito:

Las habilidades relacionadas con las TIC determinan el uso efectivo de la tecnología en la información y comunicación, aun cuando la falta de tales habilidades continúa siendo una de las barreras claves que tiene la comunidad.

Metodología de cálculo

Registra el porcentaje de personas en una población dada que tiene habilidades relacionadas con las TIC en varias áreas, diversos temas o dominios de aprendizaje, en espacios como la escuela o lugar de trabajo.

$$PICT_{3,s} = \frac{ICT_{a,s}}{P_a}$$

Donde,

$PICT_{a,s}$ Porcentaje de personas del grupo de edad (a) que tienen habilidades en las TIC (s).

$ICT_{a,s}$ Número de personas del grupo de edad (a) que tienen habilidades en las TIC (s).

P_a Población del grupo de edad (a).

Desagregación

Por edad, o por grupo de edad, sexo, ubicación y estrato socio-económico si se recopilan en la encuesta pertinente.

Fuentes:

- Encuesta de escuela o encuesta de hogares sobre el uso de las TIC.
- Se basa en una definición y metodología acordadas internacionalmente bajo la coordinación de la Unión Internacional de telecomunicaciones UIT, o en inglés (ITU).

Fuente	¿Qué es?	Disponibilidad	Formato	Observaciones
Encuesta a graduados 2014. De: MEN	Busca recoger la opinión de los graduados de la educación superior acerca de sus condiciones de ingreso al mercado laboral y su aceptabilidad, el tipo de vinculación laboral que tienen	Resultados encuestas microdatos http://www.graduadoscolombia.edu.co/html/1732/w3-article-346772.html	Excel	Acceso libre sin usuarios.
Encuesta de calidad de vida. 2015-2016. De: DANE	Las encuestas de calidad de vida surgen como respuesta a la necesidad de caracterizar la población en los diferentes aspectos involucrados en el bienestar de los hogares. Incluye un módulo de tecnologías de información y comunicación.	Acceso a microdatos del DANE. http://formularios.dane.gov.co/Anda_4_1/index.php/catalog/456/get_microdata	TXT	Se debe crear usuario, explicando el uso de los datos.
Proyecciones de población	Estas proyecciones fueron realizadas tomando como base los resultados ajustados de población del Censo 2005 y la conciliación censal 1985 - 2005; así como los análisis sobre el comportamiento de las variables determinantes de la evolución demográfica, las hipótesis y algunos comentarios sobre sus resultados	Archivo de proyecciones disponible en la página web del DANE. http://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion	Excel	Es una tabla dinámica que permite seleccionar municipios, años, sexo y genera la población estimada.

22.2. Indicador Temático 4.2

Porcentaje de jóvenes y adultos que han alcanzado al menos un nivel mínimo de dominio de competencias elementales digitales.

Observaciones:

- El nivel mínimo de competencia será medido respecto a un marco que cuenta con una métrica común para ser desarrollada.
- La alfabetización y la aritmética son importantes para usar efectivamente aplicaciones de las TIC en el manejo de información.

Propósito:

Medir la capacidad y dominio en uso de dispositivos inteligentes para manejar información, lo cual tiene importancia en diversos aspectos de la vida.

Metodología de cálculo

Refleja el porcentaje de estudiantes o personas jóvenes de un escenario relevante en educación, que han alcanzado al menos un nivel mínimo de dominio en el área temática dada.

Actuación al nivel o por encima del nivel mínimo, $PLt_{n,s,por encima del mínimo} = p$

Donde,

- | | |
|----------|---|
| p | Porcentaje de estudiantes o jóvenes en una evaluación de aprendizaje nacional o internacional, dentro de un escenario de educación |
| n | Basándose en el Estudio Internacional de Alfabetización Informática y de la Información (ICILS en inglés), el nivel de educación es el final de la secundaria inferior, sin embargo, algunos países podrían llevar a cabo la evaluación en la enseñanza secundaria superior), en la materia |
| s | en cualquier año (t-i) donde $0 \leq i \leq 5$, quien haya logrado al menos el nivel mínimo de competencia. |

Notas.

- Menor del mínimo es la proporción o porcentaje de estudiantes que no alcanzaron un estándar mínimo establecido por países de acuerdo a la definición global de competencias mínimas.

- En o por encima del mínimo es la proporción o porcentaje de estudiantes o jóvenes que han alcanzado al menos el estándar mínimo.

Desagregación

Por edad o por grupo de edad, sexo, ubicación y estrato socio-económico si se recopila en la encuesta relevante.

Fuentes:

- Encuesta de hogares.
- Se necesitan datos sobre el nivel de rendimiento que se establecen a partir de las evaluaciones nacionales e internacionales, por ejemplo, ICILS y PIAAC. Se debe garantizar que, en caso de evaluación internacional, las dos normas de evaluación midan las mismas habilidades.
- Datos de evaluaciones de aprendizaje del ministerio u organizaciones especialistas.

22.3. Indicador Temático 4.3

Tasas de escolaridad de jóvenes/adultos, por grupos de edad, actividad económica y orientación de los programas

Observaciones:

- Este indicador suele presentarse para los grupos de edad de por lo menos 25 años y mayores para asegurar que la mayoría de la población ha completado su educación.
- Los grupos de edad más jóvenes a menudo todavía están inscritos en el sistema educativo. El indicador se puede calcular para jóvenes entre 15 y 24 años si se desea. Medido a través de las evaluaciones de aprendizaje.
- Se supone que las personas con mayores niveles de rendimiento están mejor equipadas para tomar decisiones bien informadas. Por lo tanto, se supone que los altos niveles de logro en una población se correlacionan con el desarrollo sostenible.
- El indicador mide para cada nivel de educación el porcentaje de la población que completó por lo menos ese nivel de educación

Propósito: El logro educativo es una medida del capital humano de individuos y naciones enteras.

Metodología de cálculo

El indicador se calcula como el número de personas de un grupo de edad relevante que hayan completado al menos un nivel dado de educación, es expresado como el porcentaje del total de la población de la misma edad, excluyendo personas con logros educativos desconocidos.

$$EA_{nt8,AGi} = \frac{EAP_{nt8,AGi}}{P_{AGi}}$$

Donde,

$EA_{nt8,AGi}$	Porcentaje de población en el grupo de edad i que haya completado al menos el nivel de educación n , hasta e incluyendo la 'Clasificación Internacional Normalizada de la Educación CINE' (ISCED en inglés) en nivel 8 es decir nivel doctorado.
$EAP_{nt8,AGi}$	Población en el grupo de edad i que haya completado al menos el nivel de educación n , hasta e incluyendo ISCED nivel 8 (nivel doctorado).
P_{AGi}	Población en el grupo de edad i excluidas las personas con un nivel de educación desconocido.

Notas.

- El mapeo de una clasificación nacional a la ISCED no siempre es directo y puede causar dificultades entre los niveles de logro en los datos nacionales e internacionales.

Desagregación

Por edad, sexo, ubicación y estrato socio-económico, nivel de educación y otros tanto sean disponibles en encuesta o datos del censo. Las opciones de desagregación pueden estar limitadas por el tamaño de la muestra en una encuesta.

Fuentes:

- Encuesta de hogares y censos poblacionales.
- Las encuestas sobre el trabajo o según mano de obra son la fuente más común de datos sobre logros educativos.
- Poblaciones en los grupos de edad relevantes (25 años y mayores, 15-24 años, si es necesario) de acuerdo al más alto nivel de educación completado.

Fuente	¿Qué es?	Disponibilidad	Formato	Observaciones
GEIH, Gran Encuesta Integrada de Hogares. De: DANE	La encuesta se ha especializado en la medición de la estructura del mercado laboral y los ingresos de los hogares. Esta tiene una muestra total anual de 240.000 hogares aproximadamente, lo que hace que sea la de mayor cobertura a nivel nacional. Se pregunta por nivel educativo.	Microdatos de la encuesta. https://formularios.dane.gov.co/Anda_4_1/index.php/catalog/458/get_microdata	Txt	Se debe crear usuario para acceder a los datos.
Proyecciones de población	Estas proyecciones fueron realizadas tomando como base los resultados ajustados de población del Censo 2005 y la conciliación censal 1985 - 2005; así como los análisis sobre el comportamiento de las variables determinantes de la evolución demográfica, las hipótesis y algunos comentarios sobre sus resultados	Archivo de proyecciones disponible en la página web del DANE. http://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion	Excel	Es una tabla dinámica que permite seleccionar municipios, años, sexo y genera la población estimada.

23. Meta 5. Equidad

Eliminar las disparidades de género en la educación y asegurar el acceso igualitario a todos los niveles de la enseñanza y la formación profesional para las personas vulnerables, incluidas las personas con discapacidad, los pueblos indígenas y los niños en situaciones de vulnerabilidad

23.1. Indicador Mundial 5.1

Índices de paridad (sexo femenino/masculino, rurales/urbanos, quintil inferior/ superior de riqueza y otros como discapacidad, pueblos indígenas y estar afectado por un conflicto, según los datos de que se vaya disponiendo) para todos los indicadores de esta lista relativos a la educación que sea posible desglosar

Observaciones:

- Los índices de paridad representan la relación entre el valor del indicador para un grupo y el del otro.
- Típicamente, el grupo más desfavorecido es el numerador.
- Un valor igual a 1 indica la paridad entre los dos grupos.
- Cuanto más lejos está del índice de paridad, mayor es la disparidad entre los dos grupos de interés

Propósito: El indicador es una medida directa del nivel general de disparidad entre dos subpoblaciones de interés con respecto a un indicador dado

Metodología de cálculo

El valor del indicador del grupo más desfavorecido se divide por el valor del indicador de la otra subpoblación de interés.

$$DPI = \frac{[Indi]_d}{[Indi]_a}$$

Donde,

DPI	La dimensión (sexo, riqueza, ubicación, etc.) Índice de paridad.
Ind_i	El Indicador de educación 2030 (i) para el que se necesita una medida de equidad.
d	El grupo desfavorecido probable (ejemplo: mujeres, pobres, etc.)
a	El grupo favorecido probable (ejemplo: hombres, ricos, etc.)

Desagregación

Ninguna, dado que el índice de paridad compara directamente dos su-poblaciones de interés.

Fuentes:

- Se necesitan varios datos dependiendo del indicador subyacente.
- Los índices de paridad requieren datos para los grupos específicos de interés.
- Todas las fuentes utilizadas desagregadas por las variables mencionadas en el indicador.

23.2. Indicador Temático 5.2

Porcentaje de alumnos de enseñanza primaria cuya primera lengua o cuya lengua familiar es la lengua en que se imparte la enseñanza

Observaciones:

- Primera lengua o lengua familiar se define como el idioma principal de comunicación del estudiante fuera del ambiente escolar. Normalmente es el primer idioma que los estudiantes aprenden de su familia o comunidad.
- Un valor elevado del indicador indica que un gran número de alumnos de primaria se imparten en un idioma en el que son competentes, facilitando así su adaptación al entorno de aprendizaje escolar

Propósito: El indicador mide hasta qué punto los niños en la educación primaria están aprendiendo en un idioma con el que están familiarizados y en los que probablemente sean competentes.

Metodología de cálculo

El indicador se calcula como el número de pupilos en educación primaria cuya lengua familiar o primera lengua es la lengua en que se imparte la enseñanza, se expresa como un porcentaje de todos los pupilos primarios.

$$PELA_1 = \frac{EF_1}{E_1}$$

Donde,

PELA₁ Porcentaje de pupilos en educación primaria (ISCED level 1) cuya lengua familiar o primera lengua es la lengua en que se imparte la enseñanza

EF₁ Pupilos en educación primaria (ISCED level 1) cuya lengua familiar o primera lengua es la lengua en que se imparte la enseñanza.

E₁ Total de pupilos en educación primaria (ISCED level 1)

Desagregación

Por sexo de fuentes administrativas; por sexo, ubicación e ingresos según las encuestas de hogares.

Fuentes:

- Encuesta de hogares y datos administrativos.
- Número de alumnos en primaria por lengua familiar e idioma de instrucción.
- Datos administrativos de las escuelas sobre la lengua de instrucción y lenguas originales de los alumnos.
- Porcentaje de alumnos de primaria cuya lengua familiar (materna o primera) es la lengua de instrucción.

23.3. Indicador Temático 5.3

Medida en que unas políticas basadas en fórmulas explícitas reasignan recursos educativos a poblaciones desfavorecidas

Observaciones:

- El indicador incluye tanto (a) un enfoque de perspectiva a mediano plazo; y (b) a largo plazo:
 - a) El grado de compromiso de la política de financiación nacional para equiparar oportunidades de educación primaria y secundaria podría clasificarse en cuatro niveles: i) muy bajo; ii) bajos iii) medio (iv) altos.
 - b) Porcentaje del gasto público en educación asignado explícitamente a las poblaciones desfavorecidas. Dependiendo del contexto nacional dichas poblaciones pueden incluir miembros de minorías étnicas, lingüísticas y religiosas, pueblos indígenas u otros grupos.
- Es medido a través de las evaluaciones de aprendizaje.

Propósito: El indicador pretende capturar el esfuerzo que hacen los países para igualar las oportunidades de educación a través de su sistema financiero.

Metodología de cálculo

- a El indicador cualitativo derivado de documentos de políticas y / o ejercicios cualitativos como el módulo de financiamiento de la escuela del Sistema del Banco Mundial para Mejorar la Educación (SABER).
- b El indicador utiliza líneas presupuestarias detalladas para determinar el gasto público destinado a las poblaciones vulnerables. Esto requerirá una clasificación detallada de referencia de los gastos de educación y una lista convenida por grupos vulnerables.

La fórmula aún está a definirse.

Desagregación

Por poblaciones que son objeto de la fórmula de financiación (ejemplo: comunidades pobres, discapacitadas, urbanas/rurales, etc.).

Fuentes:

- Documentos sobre la política del país e informes de su presupuesto detallado para educación.
- Datos administrativos.
- Datos cualitativos sobre los sistemas educativos por parte de una organización de las Naciones Unidas, que cuente con respaldo gubernamental del marco de evaluación; y una función de los expertos para apoyar a los gobiernos en la presentación de informes.

Fuente	¿Qué es?	Disponibilidad
Ministerio de Educación Nacional, directiva ministerial 22 Ejercicios cualitativos como el módulo de finanzas escolares del Sistema del Banco Mundial para Mejorar la Educación (SABRE)	Sistema de Seguimiento a los Recursos de Gratuidad	Acceso restringido - Alcaldes y directivos de los establecimientos de educación preescolar, básica y media del país que reciben recursos destinados para la gratuidad de la educación.

23.4. Indicador Temático 5.4

Gasto en educación por estudiante y nivel de educación y fuente de financiación

Observaciones:

- La financiación inicial total proviene de fuentes gubernamentales (centrales, regionales, locales), privadas (hogares y otras fuentes privadas) e internacionales para un determinado nivel de educación (preescolar, primaria, secundaria inferior, secundaria superior, post secundaria no terciaria y terciaria educación) por alumno inscrito en ese nivel en un año dado.
- Los resultados del indicador deberían expresarse:
 - a) Como porcentaje del PIB per cápita.
 - b) En PPP \$ (constante). Paridad del Poder Adquisitivo (PPA) o en inglés Purchasing Power Parity (PPP), a menos que se proponga una desagregación adicional, este indicador considera la financiación para las instituciones públicas y privadas en conjunto.
- Usar una base por alumno es útil para la comparación, ya sea entre los niveles de educación, en el tiempo, o entre países.
- Expresar el indicador como porcentaje del PIB per cápita, o en PPA \$, también permite comparaciones entre países.

Propósito: El indicador refleja la cantidad de recursos invertidos en promedio en un solo estudiante, yendo más allá de las fuentes gubernamentales para poder calcular un costo unitario real.

Metodología de cálculo

El indicador se calcula dividiendo la financiación inicial total de distintas fuentes para un nivel de educación dado, por el número de estudiantes matriculados en ese nivel en un año dado y de nuevo dividido (i) por el PIB per cápita; y (ii) por el factor de conversión(PPP)= \$.

$$XEPGDPpc_{n,s} = \frac{XE_{n,s}}{E_n * GDPpc}$$

$$XEPPPconst_{n,s} = \frac{XE_{n,s}}{E_n * PPPconst}$$

Donde,

$XEPGDPpc_{n,s}$ Gasto por alumno en el nivel (n) de educación a partir de fuentes de financiación como porcentaje del PIB per cápita

$XEPPPconst_{n,s}$ Gasto por alumno en el nivel (n) de educación a partir de fuentes de financiación en PPP \$

$XE_{n,s}$ Gasto en el nivel (n) de educación de las fuentes de financiación

E_n	Matriculación en el nivel (n) de educación
GDPpc	PIB per cápita
PPPconst	PPP constante \$ factor de conversión

Notas.

- La parte de este indicador que se centra en el gasto público ya está disponible para un gran número de países, aunque no siempre con regularidad. La fórmula también tendría que ser ligeramente modificada si se va a utilizar la financiación inicial.

Desagregación

Por nivel de educación, fuente de financiamiento (gobierno, privado o internacional), tipo de institución (público o privado) pero con menor cobertura prevista para instituciones privadas.

Para el gasto de los hogares, también se podría calcular la desagregación por riqueza, ubicación y sexo, pero no para fuentes gubernamentales e internacionales

Fuentes:

- Datos financieros de los ministerios de finanzas y/o educación (gobierno);
- Encuestas sobre el gasto del hogar.
- Datos administrativos, como el número de estudiantes por nivel.
- Datos sobre los gastos en los niveles inferiores, donde el gobierno puede centralizarse o recaudarse directamente con las autoridades locales.

23.5. Indicador Temático 5.5

Porcentaje del total de la asistencia a la educación asignado a los países de bajos ingresos

Observaciones:

- La asistencia oficial neta ODA para el desarrollo de educación en los países de bajos ingresos, incluida la educación preescolar, primaria, secundaria y terciaria, becas y los gastos de los estudiantes en los países donantes se expresan como porcentaje de la asistencia oficial total al desarrollo.
- Los países de bajos ingresos son los definidos por (UN-OHRLLS) o en español La Oficina del Alto Representante para los Países Menos Adelantados, los Países en Desarrollo sin Litoral y los Pequeños Estados Insulares en Desarrollo, sitio oficial: (http://www.un.org/en/development/desa/policy/cdp/ldc/ldc_list.pdf)
- Un valor grande en el indicador, significa que los países de bajos ingresos están siendo priorizados para recibir ayuda en cuanto a educación.

Propósito: El indicador identifica al Official Development Assistance (ODA) o en español Asistencia Oficial para el Desarrollo (AOD) como la medida aceptada de la cooperación internacional para el desarrollo. Por lo tanto, los datos abarcan la asistencia internacional oficial a la educación.

Metodología de cálculo

El indicador refleja la ayuda total destinada a la educación de los países con menores ingresos, se expresa como el porcentaje de la ayuda a la educación.

$$PODAE_{LI} = \frac{ODAE_{LI}}{ODAE}$$

Donde,

$PODAE_{LI}$ Porcentaje de la AOD para educación asignado a los países de bajos ingresos.

$ODAE_{LI}$ Total de la AOD destinada a los países de bajos ingresos.

$ODAE$ AOD total para la educación.

Desagregación

Los datos pueden desglosarse por el país proveedor y el país receptor.

Fuentes:

- Se necesitan datos administrativos de los países donantes y otros proveedores de ayuda a la asistencia oficial para el desarrollo a la educación.
- Los datos se refieren únicamente a los flujos internacionales concesionales, proporcionados por los gobiernos.

24. Meta 6. Alfabetización y conocimientos elementales de aritmética

Asegurar que todos los jóvenes y una proporción considerable de los adultos, tanto hombres como mujeres, estén alfabetizados y tengan nociones elementales de aritmética

24.1. Indicador Mundial 6.1

Porcentaje de población de un grupo de edad determinado que alcanza al menos un nivel establecido de dominio de: a) la lectura y escritura y b) las aritméticas funcionales, por sexos

Observaciones:

- El indicador refleja el porcentaje de jóvenes (de 15 a 24 años) y de adultos (mayores de 15 años) que han alcanzado o superado un nivel dado de competencia en (a) alfabetización y (b) aritmética.
- El nivel mínimo de competencia se medirá en relación con las nuevas escalas comunes de alfabetización y cálculo aritmético actualmente en desarrollo.
- Sólo hay un umbral que divide a los jóvenes y adultos en niveles por debajo del mínimo o por encima de los niveles mínimos de competencia.
 - a) Por debajo del nivel mínimo de competencia, es el porcentaje de jóvenes y adultos que no lo han alcanzado de acuerdo con las competencias mínimas globales definidas.
 - b) En o por encima del nivel mínimo es el porcentaje de jóvenes y adultos que han alcanzado por lo menos el nivel mínimo de competencia de acuerdo con las competencias mínimas globales definidas.
- Una vez que los niveles de desempeño estén mapeados, la comunidad educativa global será capaz de identificar para cada país el porcentaje de jóvenes y adultos que lograron al menos el nivel mínimo de competencia.

Propósito: El indicador es una medida directa de los niveles de habilidad de jóvenes y adultos en las dos áreas: alfabetización y aritmética.

Metodología de cálculo

El indicador se expresa como el porcentaje de jóvenes y adultos que han alcanzado por lo menos el umbral mínimo de competencia de alfabetización y aritmética

Rendimiento alcanzado en o por encima del nivel mínimo, $PL^t a_{s, \text{por encima del mínimo}} = p$

Donde,

- p** Porcentaje de jóvenes y adultos en una evaluación nacional o transnacional de alfabetización y aritmética de adultos en el grupo de edad (**a**), en el dominio de aprendizaje en cualquier año (**t-i**) donde $0 \leq i \leq 5$, que hayan alcanzado al menos el mínimo nivel de competencia.

Desagregación

Por grupo de edad, sexo, ubicación, ingresos y tipo de habilidad.

Fuentes:

- Datos sobre el nivel de desempeño en las evaluaciones nacionales e internacionales de las habilidades en alfabetización y aritmética de adultos

Fuente	¿Qué es?	Disponibilidad	Formato	Observaciones
Examen de estado 3,5,9 y 11.	Las pruebas saber 3, 5 y 9 tienen el propósito contribuir al mejoramiento de la calidad de la educación colombiana mediante la realización de evaluaciones periódicas del desarrollo de competencias de los estudiantes de educación básica.	Bases de datos en el FTP del ICFES. ftp://ftp.icfes.gov.co/SABER11/SB11-BASES%20DE%20DATOS/	SPSS, STATA, TXT	Para ingresar al FTP se necesita: Usuario: ftpdbicfes Contraseña: dataicfes%
Resultados desde 2000. De: ICFES.	Las pruebas saber 11, pretenden: (a) seleccionar estudiantes para la educación superior, b) Monitorear la calidad de la formación que ofrecen los establecimientos de educación media y c) Producir información para la estimación del valor agregado de la educación superior.			

24.2. Indicador Temático 6.2

Tasa de alfabetización de jóvenes/adultos

Observaciones:

- Se puede decir que es el porcentaje de jóvenes (15-24 años) y adultos (15 años y más) que tienen la capacidad de leer y escribir, con comprensión, una declaración breve y sencilla sobre la vida cotidiana
- La interpretación del indicador está fuertemente ligada al método de recolección de datos.
- Las tasas de alfabetización iguales o cerca al 100% indican que casi todos los adultos o jóvenes son capaces de leer y escribir, al menos en un nivel básico.
- Algunos países calculan las tasas de alfabetización a partir de datos sobre el nivel educativo. Este enfoque no es recomendado pues las habilidades de alfabetización se pueden obtener sin participación en la educación formal.

Propósito: El indicador muestra la tasa de alfabetización, la cual indica la proporción de una población dada que tiene un nivel mínimo de habilidades de lectura y escritura.

Metodología de cálculo

El indicador excluye a las personas con un estado de alfabetización desconocido

$$LR_{AGi} = \frac{LP_{AGi}}{P_{AGi}}$$

Donde,

LR_{AGi} Tasa de alfabetización de la población en el grupo de edad i.

LP_{AGi} población alfabetizada en el grupo de edad i

P_{AGi} población en el grupo de edad i, excluidas las personas con un estado de alfabetización desconocido.

Desagregación

Por grupo de edad, sexo, ubicación y datos que se presenten en encuestas o datos del censo, como por la riqueza de los hogares, características demográficas y socioeconómicas.

Fuentes:

- Encuesta de hogares y censos poblacionales que se basan en la definición de "capaz de leer y escribir una declaración simple" de alfabetización.
- Población en el grupo de edad pertinente según el nivel de alfabetización (alfabetizado/analfabeto).

<p>GEIH, Gran Encuesta Integrada de Hogares. 2001-2017</p>	<p>La encuesta se ha especializado en la medición de la estructura del mercado laboral y los ingresos de los hogares. Esta tiene una muestra total anual de 240.000 hogares aproximadamente, lo que hace que sea la de mayor cobertura a nivel nacional. Pregunta si sabe leer o escribir.</p>	<p>Microdatos disponibles. https://formularios.dane.gov.co/Anda_4_1/index.php/catalog/458/get_microdata</p>	<p>Txt</p>	<p>Se debe crear usuario, explicando la utilidad de los datos. 2001 hasta 2017.</p>
<p>Proyecciones de población</p>	<p>Estas proyecciones fueron realizadas tomando como base los resultados ajustados de población del Censo 2005 y la conciliación censal 1985 - 2005; así como los análisis sobre el comportamiento de las variables determinantes de la evolución demográfica, las hipótesis y algunos comentarios sobre sus resultados</p>	<p>Archivo de proyecciones disponible en la página web del DANE. http://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion</p>	<p>Excel</p>	<p>Es una tabla dinámica que permite seleccionar municipios, años, sexo y genera la población estimada.</p>

24.3. Indicador Temático 6.3

Tasa de participación de jóvenes/adultos en programas de Alfabetización

Observaciones:

- El número de jóvenes (15-24 años) y adultos (mayores de 15 años) que participan en programas de alfabetización, son expresados como porcentaje de la población analfabeta de la misma edad.
- Los valores de los indicadores deben ser analizados con cautela y junto a otros indicadores que reflejen la situación de alfabetización de la población debido a sus limitaciones.
- Una tasa alta en el indicador representa un alto grado de cobertura de la población analfabeta por los programas diseñados para alcanzar ese grupo específico. El valor máximo teórico es 100%.

Propósito: Mostrar el nivel de participación de los jóvenes y adultos analfabetos en los programas de alfabetización

Metodología de cálculo

El indicador se calcula como el número de personas analfabetas en el grupo de edad pertinente, que participan en programas de alfabetización, expresados como porcentaje de la población analfabeta de la misma edad.

$$PRLP^t_a = \frac{PartLit^t_a}{IllitP^t_a}$$

Donde,

$PRLP^t_a$	Tasa de participación de la población del grupo de edad a en programas de alfabetización en el año t
$PartLit^t_a$	Participantes en los programas de alfabetización del grupo de edad a en el año t
$IllitPop^t_a$	Población analfabeta del grupo de edad a en el año t
a	15 a 24 años (jóvenes) o 15 años o más (adultos)

Desagregación

Por edad, sexo, ubicación e ingresos, dependiendo de los datos de la encuesta.

Fuentes:

- Encuesta de hogares, censos poblacionales.

- Se necesitan datos administrativos sobre la participación en programas de alfabetización para los grupos de edad definidos, combinados con estimaciones de población analfabeta para los mismos grupos de edad.

25. Meta 7. Desarrollo sostenible y ciudadanía mundial

Asegurar que todos los alumnos adquieran los conocimientos teóricos y prácticos necesarios para promover el desarrollo sostenible, entre otras cosas mediante la educación para el desarrollo sostenible y los estilos de vida sostenibles, los derechos humanos, la igualdad de género, la promoción de una cultura de paz y no violencia, la ciudadanía mundial y la valoración de la diversidad cultural y la contribución de la cultura al desarrollo sostenible

25.1. Indicador Mundial 7.1

Medida en que: i) la educación para la ciudadanía mundial y ii) la educación para el desarrollo sostenible, incluidos la igualdad de género y los derechos humanos, se generalizan a todos los niveles en: a) las políticas educativas nacionales; b) los planes de estudios; c) la formación del profesorado y d) la evaluación de los estudiantes

Observaciones:

- Está basado en la medida en que los países integran:
 - a) Global Citizenship Education (GCED), en español la Educación para la Ciudadanía Global: capacita a los estudiantes para asumir roles activos, para afrontar y resolver los desafíos globales y convertirse en contribuyentes proactivos de un mundo más pacífico, tolerante, inclusivo y seguro.
 - b) La (ESD), en español Educación para el Desarrollo Sostenible (EDS): capacita a los estudiantes para tomar decisiones informadas y acciones responsables para la integridad ambiental, la viabilidad económica y una sociedad justa para las generaciones presentes y futuras, respetando la diversidad cultural.
- Se incluye la educación sobre el cambio climático, los derechos humanos y la igualdad de género en sus sistemas educativos.
- Se podría evaluar una medida comparativa de la prioridad de GCED/ESD, como parte de uno o más temas, en relación con ciertos dominios clave de aprendizaje, como la lectura y las matemáticas.
- Medido a través de evaluaciones como, por ejemplo, si la voluntad o decisiones políticas y los recursos disponibles se han traducido en políticas concretas, programas de estudios, evaluación, etc.

Propósito: El indicador proporciona información importante sobre el nivel de compromiso nacional para alcanzar el objetivo, así como la calidad de los programas proporcionados. Además, predice la probabilidad de que se logren los resultados deseados del estudiante.

Metodología de cálculo

El método de notificación de este indicador aún no se ha definido.

Se basará en una evaluación de los informes presentados por los países que describan cómo están integrando la educación y la educación de la ciudadanía global para el desarrollo sostenible en sus políticas y sistemas de educación.

Notas:

- Este indicador puede complementarse con otros indicadores temáticos sobre GCED y EDS que propone la UNESCO, que buscan evaluar los resultados del aprendizaje de manera más directa en los dominios cognitivo, socio-emocional y de comportamiento.
- El indicador podría utilizarse para evaluar los insumos tanto en los sistemas educativos formales como no formales

Desagregación

Ninguna

Fuentes:

- Información sobre la medida en que un país dado está integrando la educación ciudadana global para el desarrollo sostenible en sus políticas y sistemas de educación
- Se necesitan informes narrativos de los países.
- Marcos de monitoreo de los derechos humanos y revisiones del sector de la educación.

Fuente	¿Qué es?	Disponibilidad
Modelo MBE 5 y 9	El método de notificación de este indicador aún no se ha definido. Se basará en una evaluación de los informes presentados por los países que describan cómo están integrando la educación y la educación de la ciudadanía global para el desarrollo sostenible en sus políticas y sistemas de educación.	NA

25.2. Indicador Temático 7.2

Porcentaje de estudiantes, por grupos de edad (o nivel de estudios), que demuestran poseer una correcta comprensión de cuestiones relativas a la ciudadanía mundial y la sostenibilidad

Observaciones:

- El GCED y la EDS abarcan todos los demás temas, incluyendo la educación sobre el cambio climático, los derechos humanos y la igualdad de género, cubiertos por el objetivo, de modo que se puede argumentar que el indicador también medirá estos ítems
- Un valor alto indica que un gran número de estudiantes, en el grupo de edad relevante, tiene por lo menos un nivel dado de comprensión en cuestiones relacionadas con la ciudadanía global y la sostenibilidad.

Propósito: El indicador es una medida directa de los resultados de aprendizaje obtenidos en la educación para la ciudadanía global, Global Equities and Commodity Derivatives (GCED) y la educación para el desarrollo sostenible (ESD) en español (EDS), crítica para la promoción del desarrollo sostenible.

Metodología de cálculo

El indicador se calcula como el número de estudiantes de una determinada edad o nivel de educación que alcanza o supera el nivel mínimo de comprensión en cuestiones relacionadas con la ciudadanía global y la sostenibilidad.

$$PEGCS_{AGi} = \frac{E_{AGi.GCS}}{P_{AGi}}$$

Donde,

PEGCS_{AGi} Porcentaje de estudiantes en el grupo de edad i con conocimiento de ciudadanía global y sostenibilidad

E_{AGi.GCS} Estudiantes en grupo de edad i con conocimiento de ciudadanía global y sostenibilidad

P_{AGi} Población en grupo de edad i

Desagregación
Por edad y sexo

Fuentes:

- Evaluaciones de aprendizaje.
- Número de estudiantes por año de edad (o nivel educativo) y su nivel de conocimiento en las áreas que se están midiendo.
- La principal fuente de datos existente para este indicador es el IEA's International Civic and Citizenship Education Study (ICCS) o en español el Estudio Internacional de Educación Cívica y Ciudadana (IICC)

Fuente	¿Qué es?	Disponibilidad	Formato	Observaciones
ICCS, Estudio Internacional de Educación Cívica y Ciudadana.	Es un programa de investigación comparativa diseñado para identificar las formas en la que los jóvenes están preparados para asumir sus roles como ciudadanos	Disponibles los microdatos http://rms.iea-dpc.org/	SPSS, SAS Y RAW	Se debe completar el formulario de permiso, se puede acceder en el mismo link.

De: Banco Interamericano de desarrollo.

25.3. Indicador Temático 7.3

Porcentaje de estudiantes de 15 años de edad que demuestran poseer un dominio suficiente de las ciencias ambientales y las ciencias de la Tierra

Observaciones:

- El indicador representa el porcentaje de estudiantes de 15 años de edad que alcanzan por lo menos un nivel mínimo de competencia en ciencias ambientales y geocientíficas (ciencias de la tierra).
- Un valor alto en el indicador, representa que un gran número de estudiantes de 15 años de edad tienen por lo menos un nivel dado de competencia y conocimiento de ciencias ambientales y geociencias
- Actualmente el indicador sólo se calcula para personas que están en la escuela. Por ende, ampliar la evaluación de las competencias a los niños y jóvenes que no asisten a la escuela requeriría nuevos tipos de encuestas que podrían ser muy costosas y difíciles de administrar.

Propósito: El indicador es una medida directa de los resultados de aprendizaje alcanzados en dos temas clave, relevantes para la promoción del desarrollo sostenible.

Metodología de cálculo

El indicador se calcula como el número de estudiantes de 15 años que alcanzan o exceden el nivel mínimo de competencia en ciencias ambientales y en geociencias, se expresa como un porcentaje sobre todos los estudiantes de 15 años de edad

$$PEESG_{15} = \frac{E_{15.ESG}}{P_{15}}$$

Donde,

PEESG₁₅	Porcentaje de estudiantes de 15 años con conocimientos de ciencias ambientales y geociencias
E_{15.ESG}	Estudiantes de 15 años con conocimientos de ciencias ambientales y geociencias
P₁₅	Población de 15 años

Desagregación

Por edad y por sexo

Fuentes:

- Encuestas de evaluación de habilidades
- Datos del número de estudiantes de 15 años y su nivel de conocimiento en las áreas que se están midiendo.

Fuente	¿Qué es?	Disponibilidad	Formato	Observaciones
Examen de estado 11. Resultados desde 2000. De: ICFES.	Las pruebas saber 11, pretenden: (a) seleccionar estudiantes para la educación superior, b) Monitorear la calidad de la formación que ofrecen los establecimientos de educación media y c) Producir información para la estimación del valor agregado de la educación superior.	Bases de datos en el FTP del ICFES. ftp://ftp.icfes.gov.co/SABER11/SB11-BASES%20DE%20DATOS/	SPSS, STATA, TXT	Para ingresar al FTP se necesita: Usuario: ftpdbicfes Contraseña: dataicfes%

Se cuenta con la prueba ciencias naturales.

25.4. Indicador Temático 7.4

Porcentaje de escuelas en las que se imparte enseñanza sobre el VIH y educación sexual basadas en aptitudes para la vida

Observaciones:

- Este indicador registra el porcentaje de escuelas que proporcionan educaciones sobre HIV o Virus de Inmunodeficiencia Humano (VIH) y sexualidad basadas en habilidades para la vida dentro del currículo formal o como parte de actividades extracurriculares.
- Refleja, además, la entrega del plan de estudios en apoyo de los programas nacionales de prevención del VIH
- Un valor alto en el indicador expresa que un gran número de escuelas en el nivel dado de educación, proveen educación acerca del VIH y sexualidad, basada en habilidades para la vida a los estudiantes.

Propósito: El indicador permite evaluar los progresos realizados en la aplicación de la educación sobre el VIH y la sexualidad basada en las aptitudes para la vida en todas las escuelas.

Metodología de cálculo

El indicador se calcula como el número de escuelas en cada nivel de educación que proporciona educación sobre VIH y sexualidad basadas en habilidades para la vida, se expresa como un porcentaje de todas las escuelas en el nivel dado de educación

$$PSHIV_n = \frac{SHIV_n}{S_n}$$

Donde,

PSHIV_n Porcentaje de escuelas en el nivel n de educación que proporciona educación sobre el VIH y la sexualidad basadas en habilidades para la vida.

$SHIV_n$ Escuelas de nivel de educación n que proporcionan educación sobre el VIH y la sexualidad basadas en habilidades para la vida

S_n Número total de escuelas en el nivel de educación n.

Desagregación

Por nivel de educación

Fuentes:

- Encuesta de hogares.
- Se necesitan datos administrativos de las escuelas y otros proveedores de educación y formación.
- Número de escuelas en cada nivel de educación que proporciona educación sobre el VIH y la sexualidad y el número total de escuelas en el mismo nivel.

25.5. Indicador Temático 7.5

Medida en que se aplica en el país el marco del Programa Mundial para la educación en derechos humanos (de conformidad con la resolución 59/113 de la Asamblea General de las Naciones Unidas)

Observaciones:

- Se trata de medir la cantidad y la calidad de las acciones de los países y su compromiso con la incorporación de la educación en derechos humanos.
- El plan de acción para el período 2015-2019 se centra en:
 - a) consolidar las acciones en las dos etapas anteriores: educación en derechos humanos en las escuelas primarias y secundarias (2005-2009); y la educación en derechos humanos para la educación superior y programas de capacitación en derechos humanos para maestros y educadores, funcionarios públicos, funcionarios encargados de hacer cumplir la ley y personal militar (2010-2014).
 - b) promover la formación en derechos humanos para profesionales de los medios de comunicación y periodistas.

Propósito: El indicador es una medida del compromiso del gobierno al asegurar que los estudiantes, en todos los niveles de la educación, tengan la oportunidad de adquirir conocimientos y habilidades requeridos en el área de los derechos humanos para promover el desarrollo sostenible.

Metodología de cálculo

El método de notificación de este indicador aún no se ha definido.

Se basará en una evaluación de los informes presentados por los países que describan cómo están implementando el World Programme for Human Rights Education (OHCHR) o Programa Mundial de Educación en Derechos Humanos (PMEDH)

Desagregación

Ninguna

Fuentes:

- Se necesita información sobre la medida en que un país dado está aplicando el OHCHR. El formato exacto de los informes aún no se ha definido.
- Datos administrativos

Meta 7A. Instalaciones escolares y entornos de aprendizaje

Construir y adecuar instalaciones educativas que tengan en cuenta las necesidades de los niños y las personas con discapacidad y las diferencias de género, y que ofrezcan entornos de aprendizaje seguros, no violentos, inclusivos y eficaces para todos

25.6. Indicador Mundial 7A.1

Porcentaje de escuelas con acceso a: a) electricidad, b) Internet con fines pedagógicos, c) computadoras para fines pedagógicos, d) infraestructura adaptada y materiales adaptados para alumnos con discapacidades, e) suministro básico de agua potable, f) instalaciones sanitarias básicas para cada sexo, y g) instalaciones elementales para lavarse las manos (con arreglo a las definiciones de agua, saneamiento e higiene)

Observaciones:

- Se refiere al porcentaje de escuelas por nivel de educación con acceso a la instalación o servicio dado que permiten el uso adecuado y sostenible de la infraestructura de TIC con fines educativos.
- Tiene en cuenta el uso de computador y de internet con fines pedagógicos. En cuanto al primero, puede ser una computadora de escritorio, una computadora portátil o netbooks, pero no tabletas o dispositivos portátiles similares. Respecto al acceso a internet, puede realizarse a través de una banda fija, estrecha, ancha fija o de una red móvil.
- La infraestructura adaptada se define como cualquier entorno construido relacionado con instalaciones educativas que sean accesibles a todos los usuarios, incluyendo aquellos con diferentes tipos de discapacidad. Espacios de saneamiento básico, de lavado de manos y por ende con servicio de agua potable.
- Los materiales adaptados incluyen materiales de aprendizaje como libros de texto, materiales de instrucción, evaluaciones y otros que están disponibles y proporcionados en formatos apropiados como audio, braille, lenguaje de señas y formatos simplificados que permiten a estudiantes y maestros con discapacidades funcionales acceder al aprendizaje y participar plenamente en el ambiente escolar.
- Un valor alto en el indicador, refleja que las escuelas tienen buen acceso a los servicios e instalaciones pertinentes. Idealmente cada escuela debe tener acceso a todos estos servicios e instalaciones.

Propósito: El indicador mide el acceso en las escuelas a los servicios básicos e instalaciones necesarias para asegurar un ambiente de aprendizaje seguro y efectivo para todos los estudiantes.

Metodología de cálculo

El indicador se calcula como el número de escuelas en un nivel dado de educación con acceso a las instalaciones pertinentes, se expresa como un porcentaje de todas las escuelas en ese nivel de educación.

$$PS_{n,f} = \frac{S_{n,f}}{S_n}$$

Donde,

$PS_{n,f}$ Porcentaje de escuelas en el nivel n de educación con acceso a las instalaciones f

$S_{n,f}$ Escuelas de nivel n de educación con acceso a las instalaciones f

S_n Número total de escuelas en el nivel de educación n.

Desagregación

Por nivel de educación

Fuentes:

- Se necesitan datos sobre el número de escuelas en cada nivel de educación con y sin acceso a las instalaciones dadas.
- Datos administrativos de las escuelas y otros proveedores de educación o capacitación.

Fuente	¿Qué es?	Disponibilidad	Formato	Observaciones
Investigación de Educación Formal 2007 - 2016. De: DANE	Censo de establecimientos educativos que ofrecen educación básica formal en los niveles educativos de preescolar, básica primaria, básica secundaria y media.	Acceso a microdatos del DANE. http://formularios.dane.gov.co/Anda_4_1/index.php/catalog/503/get_microdata	TXT	Se debe crear usuario, explicando el uso de los datos.

25.7. Indicador Temático 7A.2

Porcentaje de estudiantes que son objeto de matoneo, castigos corporales, acoso, violencia, discriminación sexual y abusos

Observaciones:

- Puede haber casos en que algunos estudiantes se sienten intimidados de tal manera que no informan de incidentes violentos vividos, lo que resulta en una sobreestimación de la seguridad del entorno escolar.
- Un valor alto en el indicador, significa que un gran número de estudiantes en el nivel de educación dado están experimentando violencia y abuso en la escuela y por lo tanto que la escuela no es un ambiente seguro en el que promover el aprendizaje.

Propósito: El indicador proporciona información sobre el alcance de la violencia auto informada y el acoso en las escuelas

Metodología de cálculo

El indicador se calcula como el número de estudiantes en un determinado nivel que informan haber experimentado cualquier tipo de violencia, se expresa como un porcentaje de todos los estudiantes en el mismo nivel de educación

$$PEB_n = \frac{EB_n}{E_n}$$

Donde,

PEB_n Porcentaje de estudiantes en el nivel de educación n que experimentan intimidación, castigo corporal, acoso, violencia, discriminación o abuso sexual

EB_n Estudiantes en el nivel de educación n que experimentan intimidación, castigo corporal, acoso, violencia, discriminación o abuso sexual

E_n Matriculación total en el nivel de educación n

Desagregación

Por sexo y nivel de educación

Fuentes:

- Encuesta en las escuelas.
- Se necesitan datos sobre el número de estudiantes en cada nivel de educación que informan que han experimentado abuso y el número total de estudiantes en el mismo nivel.
- Una opción es el Global School-based Student Health Survey (GSHS), en español, la Encuesta Mundial de Salud a Escolares

Fuente	¿Qué es?	Disponibilidad	Formato	Observaciones
GSHS. Encuesta Mundial de Salud a Escolares . De: World Health Organization	Censo de establecimientos educativos que ofrecen educación básica formal en los niveles educativos de preescolar, básica primaria, básica secundaria y media.	Acceso a microdatos. http://www.who.int/chp/gshs/colombiadataset/en/	TXT	Acceso libre

25.8. Indicador Temático 7A.3

Número de agresiones contra estudiantes, personal de educación e instituciones educativas

Observaciones:

- Medido por ataques dirigidos contra estudiantes, maestros y otro personal o contra edificios, materiales e instalaciones, incluido el transporte escolar. Estos ataques pueden haberse realizado por razones políticas, militares, ideológicas, sectarias, étnicas o religiosas
- En la actualidad, los datos disponibles para el seguimiento global se recopilan a partir de informes de cada país por una amplia variedad de partes interesadas, que se utilizan para calcular los tipos de riesgos de seguridad operacional y seguir las tendencias.
- Un valor alto indica un gran número de ataques a escuelas, por lo tanto, que la escuela no es un entorno seguro en el que promover el aprendizaje

Propósito: El indicador es una medida amplia de la seguridad de los entornos de aprendizaje.

Metodología de cálculo

El método de reporte de este indicador aún no se ha definido.

Se basará en la evaluación de una amplia gama de informes de países sobre ataques de diversos tipos en escuelas y otros centros de aprendizaje

Desagregación

Por tipo de ataque y por la ubicación determinada

Fuentes:

- Información sobre el número y tipos de ataques contra escuelas y otros centros de aprendizaje a partir de múltiples fuentes, incluyendo organismos gubernamentales, organizaciones de derechos humanos y desarrollo, ONGs, organizaciones de rescate académico, sindicatos, agencias de la ONU e informes de los medios de comunicación.

26. Meta 7B. Becas

De aquí a 2020, aumentar considerablemente a nivel mundial el número de becas disponibles para los países en desarrollo, en particular los países menos adelantados, los pequeños Estados insulares en desarrollo y los países africanos, a fin de que sus estudiantes puedan matricularse en programas de enseñanza superior, incluidos programas de formación profesional y programas técnicos, científicos, de ingeniería y de tecnología de la información y las comunicaciones, de países desarrollados y otros países en desarrollo.

26.1. Indicador Mundial 7B.1

Volumen de las corrientes de asistencia oficial para el desarrollo destinado a becas, por sectores y tipos de estudios

Observaciones:

- En general, se carecen de datos detallados y comparables internacionalmente sobre las becas para países en desarrollo que ofrecen universidades, colegios, fundaciones, ONG y otras fuentes.
- Un valor alto indica que hay un mayor gasto en estudiantes del país beneficiario dado, para estudiar en el extranjero. No indica el número de estudiantes que reciben apoyo.

Propósito: La Asistencia Oficial neta para el Desarrollo (AOD) u (*ODA*, por sus *siglas* en inglés) es la medida aceptada de la cooperación internacional para el desarrollo. Por lo tanto, los datos abarcan la asistencia oficial internacional para proporcionar lugares de educación e otros países como parte del desarrollo de estos mismos.

Metodología de cálculo

El indicador se calcula como la suma de la ayuda oficial al desarrollo total para becas en estudio en el exterior por sector y tipo de estudio otorgado a los estudiantes del país beneficiario, es expresado en dólares estadounidenses.

Desagregación

Por proveedor de ayuda

Fuentes:

- Se necesitan datos administrativos de los países donantes y otros proveedores de ayuda para la ayuda oficial al desarrollo neto a la educación

26.2. Indicador Temático 7B.2

Número de becas concedidas para cursar estudios superiores, por países beneficiarios

Observaciones:

- Por lo general, los países beneficiarios no tendrán acceso a todos los datos sobre las becas para estudios en el extranjero otorgadas a sus habitantes.
- En la mayoría de los países en los que estudian estos estudiantes, hay una fuente de información sobre becas otorgadas a estudiantes extranjeros, ya que pueden ser ofrecidos por diferentes fuentes.
- Puede haber problemas con la identificación de los países de origen de los estudiantes.
- Un valor alto no indica la cantidad de apoyo financiero ni si esto es suficiente para cubrir todos los costos de los estudiantes relacionados con su estudio.

Propósito: El indicador es una medida directa de las becas para estudiar en el extranjero según se define en el objetivo

Metodología de cálculo

El indicador se calcula como la suma de todas las becas concedidas en un año académico dado por los países donantes o de acogida a los estudiantes del país beneficiario dado para estudiar en el extranjero.

Desagregación

Por sexo a determinar

Fuentes:

- Se necesita el número de becas para educación superior otorgadas por países donantes o de acogida a estudiantes de países beneficiarios para estudiar en el extranjero.
- Datos administrativos de los proveedores de becas de educación superior y de las instituciones de educación superior beneficiarias.

27. Meta 7C. Docentes

Aumentar considerablemente la oferta de docentes calificados, incluso mediante la cooperación internacional para la formación de docentes en los países en desarrollo, especialmente los países menos adelantados y los pequeños Estados insulares en desarrollo

27.1. Indicador Mundial 7C.1

Proporción de docentes de: a) enseñanza preescolar; b) primaria; c) primer ciclo de secundaria y d) segundo ciclo de secundaria que han recibido al menos la formación de profesores organizada mínima (por ejemplo, en didáctica) antes de la entrada en funciones y en el empleo exigida para enseñar en el nivel de que se trate en un país dado

Observaciones:

- Lo ideal sería que todos los profesores recibieran una formación pedagógica adecuada, apropiada y pertinente para enseñar en el nivel de educación escogido y ser académicamente bien calificados en los temas que se espera que enseñen.

- El indicador debe calcularse por separado para las instituciones públicas y privadas.
- Los requisitos nacionales mínimos de formación pueden variar ampliamente de un país a otro. Esta variabilidad entre países disminuye la utilidad del seguimiento global, pues el indicador sólo mostraría el porcentaje de logro de las normas nacionales, no si los profesores de diferentes países tienen niveles similares de formación.

Propósito: El indicador mide la proporción de la mano de obra docente que está pedagógicamente bien entrenada.

Metodología de cálculo

El indicador se calcula como el número de maestros en un determinado nivel de educación que se capacita, se expresa como un porcentaje de todos los maestros en ese nivel de educación.

$$PTT_n = \frac{TT_n}{T_n}$$

Donde,

PTT_n Porcentaje de maestros capacitados en el nivel de educación n

TT_n Maestros capacitados en el nivel de educación n

T_n Total de maestros de nivel de educación n

n 02 (preescolar), 1 (primaria), 2 (secundaria inferior), 3 (secundaria superior) y 23 (secundaria)

Desagregación

Por sexo, nivel de educación y tipo de institución educativa (pública o privada)

Fuentes:

- Datos administrativos de las escuelas y otros centros de aprendizaje organizados.
- Se necesitan datos sobre el número de maestros en cada nivel de educación que están formados y número total de maestros en cada nivel.

Fuente	¿Qué es?	Disponibilidad	Formato	Observaciones
Investigación de Educación Formal 2007 - 2016.	Censo de establecimientos educativos que ofrecen educación básica formal en los niveles educativos de preescolar, básica primaria, básica secundaria y media.	Acceso a microdatos del DANE.	TXT	Se debe crear usuario, explicando el uso de los datos.
De: DANE		http://formularios.dane.gov.co/Anda_4_1/index.php/catalog/503/get_microdata		

27.2. Indicador Temático 7C.2

Porcentaje de profesores cualificados con arreglo a las normas nacionales, por nivel de educación y tipo de institución

Observaciones:

- Lo ideal es que todos los profesores reciban una formación pedagógica adecuada, apropiada y pertinente para enseñar en el nivel de educación elegido y ser académicamente bien calificados en los temas que se espera que enseñen.
- El indicador debe calcularse por separado para las instituciones públicas y privadas.
- Los requisitos de calificaciones académicas nacionales pueden variar de un país a otro. Esta variabilidad entre los países disminuye la utilidad del seguimiento global, ya que el indicador solo mostraría el porcentaje de logro de los estándares nacionales, más no si los profesores tienen niveles similares de calificaciones académicas en diferentes países.

Propósito: El indicador mide la proporción de la mano de obra docente académicamente bien calificada.

Metodología de cálculo

El indicador se calcula como el número de profesores calificados en un nivel de educación determinado, se expresa como un porcentaje de todos los maestros en ese nivel de educación

$$PQT_n = \frac{QT_n}{T_n}$$

Donde,

PQT_n Porcentaje de profesores calificados en el nivel de educación n

QT_n Profesores calificados en el nivel n de educación

T_n Total de profesores de nivel de educación n

n 02 (pre-primaria), 1 (primaria), 2 (secundaria inferior), 3 (secundaria superior) y 23 (secundaria)

Desagregación

Por sexo, nivel de educación y tipo de institución (pública o privada)

Fuentes:

- Se necesita el número de profesores calificados en cada nivel de educación y el número total de maestros en cada nivel.

- Datos administrativos de las escuelas y otros centros de aprendizaje organizados

Fuente	¿Qué es?	Disponibilidad	Formato	Observaciones
Investigación de Educación Formal 2007 - 2016. De: DANE	Censo de establecimientos educativos que ofrecen educación básica formal en los niveles educativos de preescolar, básica primaria, básica secundaria y media. Docentes en educación básica.	Acceso a microdatos del DANE. http://formularios.dane.gov.co/Anda_4_1/index.php/catalog/503/g-et_microdata	TXT	Se debe crear usuario, explicando el uso de los datos.
SNIES. Sistema Nacional de Información de la Educación Superior. 2014 - 2016. De: MEN	Es un sistema de información que ha sido creado para responder a las necesidades de información de la educación superior en Colombia. Docentes en educación superior.	Acceso a agregados por institución. http://www.mineducacion.gov.co/sistemasdeinformacion/1735/w3-article-212400.html	Excel	Acceso libre sin usuarios.

27.3. Indicador Temático 7C.3

Proporción alumnos/profesores cualificados, por nivel de educación

Observaciones:

- Dado que los maestros cualificados desempeñan un papel clave en la calidad de la educación siempre y cuando la proporción alumno / maestro cualificado, se considera determinante en los resultados del aprendizaje y un indicador de la calidad general de un sistema educativo.
- Un maestro cualificado es aquel que tiene por lo menos las calificaciones académicas mínimas requeridas para enseñar sus materias en el nivel relevante en un país dado.
- Las proporciones de alumnos/profesores cualificados "ideales" pueden depender de una amplia variedad de factores complejos, incluyendo la edad, las necesidades académicas de los alumnos representados, la experiencia, habilidad y efectividad de los profesores.
- Las proporciones alumno/profesor no son equivalentes al tamaño promedio de la clase. Por lo que los requisitos nacionales de formación de profesores pueden variar de un país a otro.

Propósito: El indicador mide la carga de trabajo de los maestros cualificados y la asignación de recursos humanos en las instituciones educativas, da una indicación general del promedio de tiempo y atención individual que un alumno recibirá de profesores capacitados.

Metodología de cálculo

El indicador se calcula como el número total de alumnos y estudiantes en el nivel correspondiente se divide por el número de profesores cualificados en el mismo nivel.

$$PQTR_n = \frac{E_n}{QT_n}$$

Donde,

$PQTR_n$ Proporción de maestros calificados por alumnos en el nivel de educación n

E_n Alumnos matriculados en el nivel de educación n

QT_n Profesores calificados en el nivel n de educación

n 02 (pre-primaria), 1 (primaria), 2 (secundaria inferior), 3 (secundaria superior) y 23 (secundaria)

Desagregación

Por nivel de educación y por tipo de institución (pública o privada)

Fuentes:

- Se necesitan datos sobre el número de alumnos y profesores cualificados en cada nivel de educación.
- Datos administrativos de las escuelas y otros centros de aprendizaje organizados

Fuente	¿Qué es?	Disponibilidad	Formato	Observaciones
Investigación de Educación Formal 2007 - 2016. De: DANE	Censo de establecimientos educativos que ofrecen educación básica formal en los niveles educativos de preescolar, básica primaria, básica secundaria y media. Docentes en educación básica.	Acceso a microdatos del DANE. http://formularios.dane.gov.co/Anda_4_1/index.php/catalog/503/get_microdata	TXT	Se debe crear usuario, explicando el uso de los datos.
SNIES. Nacional de Información de Educación Superior. 2014 - 2016. De: MEN	Es un sistema de información que ha sido creado para responder a las necesidades de información de la educación superior en Colombia. Docentes en educación superior. Estudiantes matriculado	Acceso a agregados por institución. http://www.mineduccion.gov.co/sistemasdeinformacion/1735/w3-article-212400.html	Excel	Acceso libre sin usuarios.
SIMAT, integrado de matriculados De: Ministerio de Educación Nacional de Colombia	Es una herramienta que permite organizar y controlar el proceso de matrícula en todas sus etapas, así como tener una fuente de información confiable y disponible para la toma de decisiones.	Restringido	No aplica	

27.4. Indicador Temático 7C.4

Proporción de alumnos/profesores formados, por nivel de educación y tipo de institución

Observaciones:

- La proporción alumno/docente capacitado se considera un determinante en los resultados del aprendizaje y un indicador de la calidad general de un sistema educativo.
- Un maestro capacitado es aquel que ha recibido por lo menos el mínimo de formación pedagógica organizada a un nivel de profesorado al nivel pertinente en un país determinado.
- Los resultados del indicador pueden compararse con las normas nacionales establecidas sobre el número de alumnos por maestro capacitado para cada nivel de educación.
- Se debe tener en cuenta la existencia de la enseñanza a tiempo parcial, las jornadas escolares, las clases multigrado y otras prácticas que pueden afectar la precisión y la significación de las proporciones estudiante/profesor.

Propósito: El indicador mide la carga de trabajo de los maestros capacitados y la asignación de recursos humanos en las instituciones educativas, y dar una indicación general del promedio de tiempo y atención individual que un alumno recibirá de maestros capacitados.

Metodología de cálculo

El indicador se calcula como el número total de estudiantes en el nivel pertinente se divide por el número de profesores capacitados en el mismo nivel

$$PTTR_n = \frac{E_n}{TT_n}$$

Donde,

PTTR_n Proporción de maestros capacitados para alumnos en el nivel de educación n

E_n Alumnos matriculados en el nivel de educación n

TT_n Profesores o maestros capacitados en el nivel de educación n

n 02 (pre-primaria), 1 (primaria), 2 (secundaria inferior), 3 (secundaria superior) y 23 (secundaria)

Desagregación

Por nivel de educación y tipo de institución (pública o privada)

Fuentes:

- Se necesitan datos sobre el número de alumnos y profesores capacitados en cada nivel de educación.
- Datos administrativos de las escuelas y otros centros de aprendizaje organizados

Fuente	¿Qué es?	Disponibilidad	Formato	Observaciones
Investigación de Educación Formal 2007 - 2016. De: DANE	Censo de establecimientos educativos que ofrecen educación básica formal en los niveles educativos de preescolar, básica primaria, básica secundaria y media. Docentes en educación básica.	Acceso a microdatos del DANE. http://formularios.dane.gov.co/Anda_4_1/index.php/catalog/503/g-et_microdata	TXT	Se debe crear usuario, explicando el uso de los datos.
SNIES. Sistema Nacional de Información de la Educación Superior. 2014 - 2016. De: MEN	Es un sistema de información que ha sido creado para responder a las necesidades de información de la educación superior en Colombia. Docentes en educación superior. Estudiantes matriculado	Acceso a agregados por institución. http://www.mineduccion.gov.co/sistemasdeinformacion/1735/w3-article-212400.html	Excel	Acceso libre sin usuarios.
SIMAT, integrado de matriculados De: Ministerio de Educación Nacional de Colombia	Es una herramienta que permite organizar y controlar el proceso de matrícula en todas sus etapas, así como tener una fuente de información confiable y disponible para la toma de decisiones.	Restringido	No aplica	

27.5. Indicador Temático 7C.5

Salario en promedio de los docentes comparado con el de otras profesiones que exigen un nivel comparable de calificación escolar

Observaciones:

- La razón de este indicador es que, si los salarios en la profesión docente son atractivos, es más probable que atraigan candidatos de calidad.
- El sueldo estatutario inicial (sin incluir bonificaciones, subsidios y cotizaciones a la pensión) se recomienda como una mejor opción que el vago o vano salario "medio".
- Se sugiere decidir sobre algunas profesiones (entre 4 a 5) que, en general, requieren un nivel similar de calificación para un maestro, y recopilar datos de sueldos sobre éstas de país a país.

Propósito: El indicador da una idea del atractivo de la profesión como docente en comparación a otras profesiones que requieren un nivel de cualificación similar.

Metodología de cálculo

El indicador se calcula como el sueldo bruto anual estatutario para un profesor calificado de primaria o secundaria en instituciones públicas, dividido por el salario inicial bruto de una canasta de profesiones que requieren un nivel de educación comparable.

$$RTS_n = \frac{TS_n}{OS_n}$$

Donde,

RTS_n Promedio del salario inicial estatutario del profesor en el nivel de educación n en relación con otras profesiones

TS_n Salario inicial bruto anual de un profesor calificado para el nivel de educación n

OS_n Sueldo base promedio bruto anual de la canasta de profesiones que requieren un nivel similar de calificaciones

n 1 (primaria) o 23 (secundaria)

Desagregación, Por nivel de educación

Fuentes:

- Se necesitan escalas de sueldos para maestros calificados en escuelas públicas de educación primaria y secundaria.
- Escalas salariales de las profesiones que requieren un nivel similar de calificaciones.

- Las encuestas de población activa o encuestas socioeconómicas llevadas a cabo por las oficinas de estadística pueden recopilar información sobre la ocupación y los salarios.

Fuente	¿Qué es?	Disponibilidad	Formato	Observaciones
Observatorio laboral para la educación. 2009-2015 De: MEN	Censo de establecimientos educativos que ofrecen educación básica formal en los niveles educativos de preescolar, básica primaria, básica secundaria y media. Docentes en educación básica.	Estadísticas agregadas http://bi.mineducacion.gov.co:8380/eportal/web/men-observatorio-laboral/ubicacion-geografica	Online	Se debe manejar a plataforma proporcionada por el MEN para realizar estos cálculos.

27.6. Indicador Temático 7C.6

Tasa de deserción de profesores, por niveles de educación

Observaciones:

- Hace alusión a que la escasez de maestros es un factor que contribuye significativamente a ampliar las brechas de equidad en el acceso a la educación y el aprendizaje.
- Este indicador es el porcentaje de maestros de un determinado nivel de educación que abandonan la profesión en un año escolar determinado. No proporciona información sobre las razones por las cuales los profesores abandonan la profesión.
- Se debe evitar el doble conteo con respecto a maestros que enseñan más de un nivel de educación. Asimismo, debe tenerse en cuenta la existencia de enseñanza a tiempo parcial, turnos escolares, clases multigrado y otras prácticas que puedan afectar la precisión del número de profesores y de los nuevos participantes en la profesión docente.

Propósito: El indicador evalúa y vigila la deserción de maestros es esencial para asegurar un suministro suficiente de maestros calificados y bien capacitados, así como para su despliegue, apoyo y manejo efectivos.

Metodología de cálculo

El indicador, refleja la tasa de deserción, es el número de egresados expresados como porcentaje del número total de maestros en el año t-1.

$$TAR_{n,t} = \frac{(T_{n,t} - T_{n,t-1}) + NET_{n,t}}{T_{n,t-1}}$$

Donde,

TAR_{n,t} Tasa de deserción de maestros del nivel de educación n en el año t

T_{n,t} Maestros del nivel de educación n en el año t

$T_{n,t-1}$ Maestros del nivel de educación n en el año t-1

$NET_{n,t}$ Nuevos Maestros del nivel de educación n en el año t
n 02 (educación preescolar), 1 (educación primaria), 2 (educación secundaria inferior), 3 (educación secundaria superior) y 23 (educación secundaria)

Desagregación, Por sexo y nivel de educación

Fuentes:

- Se necesitan datos sobre el número de maestros en cada nivel de educación en los años t y t-1 y el número de maestros nuevos en cada nivel en el año t.
- Datos administrativos de las escuelas y registros de recursos humanos del personal educativo.

27.7. Indicador Temático 7C.7

Porcentaje de docentes que recibieron formación en el empleo en los últimos 12 meses, por tipos de formación

Observaciones:

- Se refiere a que los programas de formación de maestros en servicio suelen apuntar a mejorar la calidad de la instrucción en el aula.
- Además de los requisitos de su formación antes y durante un empleo, los profesores deben recibir de vez en cuando una formación pertinente en el servicio para el nivel de educación que enseñan a fin de mejorar su dominio de la enseñanza.
- Es el porcentaje de docentes por nivel educativo que, durante el último curso académico, han recibido una formación en el puesto de trabajo, necesaria para la enseñanza a nivel pertinente en un país, por tipo de la formación recibida.

Propósito: El indicador mide la proporción de la fuerza de trabajo docente que recibió formación durante el último año académico

Metodología de cálculo

El indicador se calcula como el número de maestros de un determinado nivel de educación que recibieron, en el último año, formación en un determinado servicio, se expresa como un porcentaje de todos los maestros en ese nivel de educación

$$PTIN_{n,j} = \frac{T_{n,j}}{T_n}$$

Donde,

$PTIN_{n,j}$ Porcentaje de docentes en el nivel de educación n que reciben formación en el último año del tipo j

$T_{n,j}$ Profesores del nivel de educación n que reciben formación en el último año del tipo j

T_n Total de maestros en el nivel de educación n.

n 02 (educación preescolar), 1 (educación primaria), 2 (educación secundaria inferior), 3 (educación secundaria superior) y 23 (educación secundaria)

Desagregación

Por sexo, nivel de educación y tipo de formación.

Fuentes:

- Datos administrativos o encuestas escolares.

- Se necesitan datos sobre el número de maestros en cada nivel de educación, que recibieron capacitación en servicio de cada tipo en el último año y sobre el número total de maestros en cada nivel de educación.
- Encuestas de directores o datos administrativos de escuelas y otros centros de aprendizaje organizados nacionales de formación docente.

Flujo de Validación de Datos

28. Descripción del indicador

Este proceso se encarga de detallar la información relevante de cada indicador que se propone por la UNESCO, donde se identifica a que hace referencia el indicador una descripción de este y las limitaciones que se pueden tener en cuanto a lo relacionado con el indicador.

Documento donde se plasma: Documento de caracterización

29. Identificar variables necesarias para calcular el indicador

Se ajusta el documento de caracterización de indicadores con la descripción de cada variable que conforma el indicador, indicando a que hace referencia la variable.

Documento donde se plasma: Documento de caracterización

30. Identificar fuente que puede proveer la información

En este proceso se identifica la fuente de los datos, el tipo de acceso, la disponibilidad de los datos.

Documento donde se plasma: Listado de variables y las fuentes que la proveen

31. Validación de la calidad de los datos

Este gran proceso inicia con la validación del tipo de fuente o dato relacionado, que puede ser:

- Documento oficial, decretos o normas que están alineadas con lo propuesto por la UNESCO
- Información calculada previamente, por ejemplo, un índice nacional ya conocido y calculado por instituciones oficiales o privadas de reconocimiento nacional.
- Conjunto de datos, set de datos que contienen información que permite el cálculo o el acercamiento a la variable o indicador deseado.

Documento donde se plasma: Documento de caracterización

31.1. Verificación de la pertinencia del documento con la necesidad

Este proceso se activa en el caso que la fuente para ser evaluada corresponda con un documento, y consiste en hacer la evaluación de relevancia con el indicador y las variables relacionadas.

31.2. Verificación de frecuencia de actualización

Este proceso se invoca en la validación de un set de datos con el propósito de documentar que tal actualizado está el dato y la expectativa de actualización que pueda llegar a tener.

31.3. Verificación de la completitud de los datos por variable

Este proceso se aplicará en el caso que se haga referencia a un conjunto de datos, se debe verificar la completitud relevante para los análisis que se requieran, es decir que, para la mayoría de registros, superior al 70%, exista un valor asociado.

31.4. Validar la fuente

Este proceso se utiliza para las fuentes calculadas previamente donde la revisión que se hace es sobre el ente que calcula el valor, por la reputación y el reconocimiento nacional e internación que tenga en el campo asociado a la variable o indicador propuesto.

31.5. Consolidar proceso de evaluación

Se debe recopilar los resultados obtenidos de los procesos ya cursados, para los diferentes tipos de fuentes, para que con esta información se pueda generar un insumo de la decisión de utilización o no.

Documento donde se plasma: Documento de caracterización

32. Utilizar el indicador

Este proceso se activa con la recepción del concepto resultante de consolidar el proceso de evaluación de calidad de los datos y que haya sido positivo en el proceso asociado a la revisión, para su uso según la propuesta de indicador, y este resultado permitirá identificar si es o no posible el cálculo necesario para responder al indicador o variable que se está evaluando.

33. Seleccionar Fuente

Este proceso verifica que la fuente sea seleccionable porque satisface las necesidades identificadas por el indicador.

34. Aplicar Directrices del gobierno de datos

Este paso da inicio a la aplicación del gobierno de datos, que implica la aplicación del documento del mismo nombre y que está enmarcado en el observatorio de datos.

35. Buscar fuente de información alterna o substituta

En los casos que la información no satisfaga alguno de los procesos, será necesario identificar si información que pueda "cubrir" la necesidad planteada por el indicador.

Documento donde se plasma: Documento de Propuestas de Información Alterna y Recomendación.

36. Solicitar información

Se activa una vez se identifique que existe información que podría satisfacer las demandas del indicador o la variable, y a la cual no se tiene acceso.

Documento donde se plasma: Documento de Propuestas de Información Alterna y Recomendación.

Ilustración 5. Flujos del proceso de validación

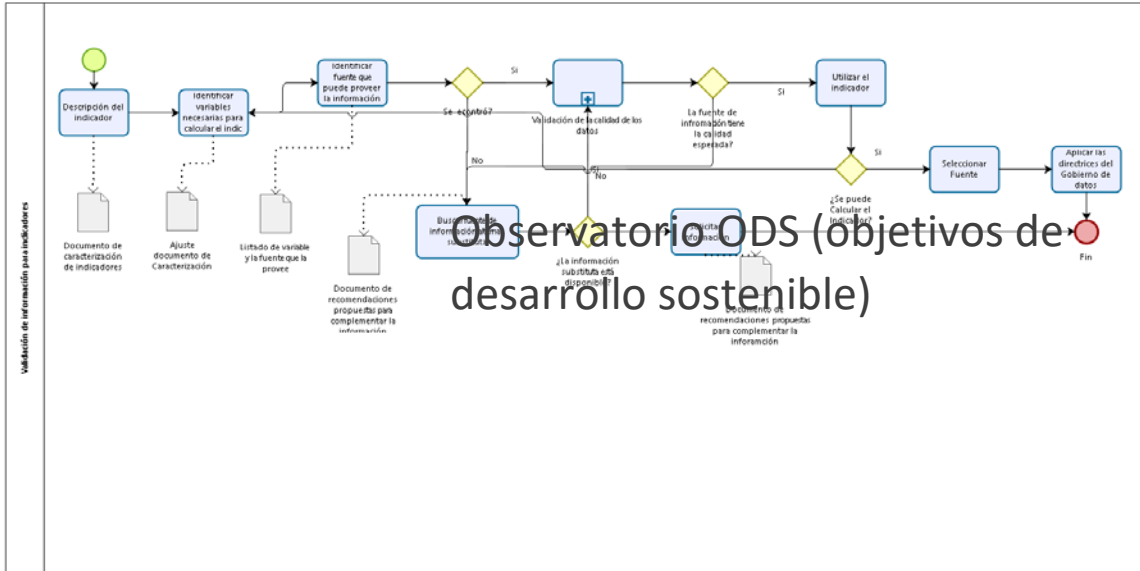
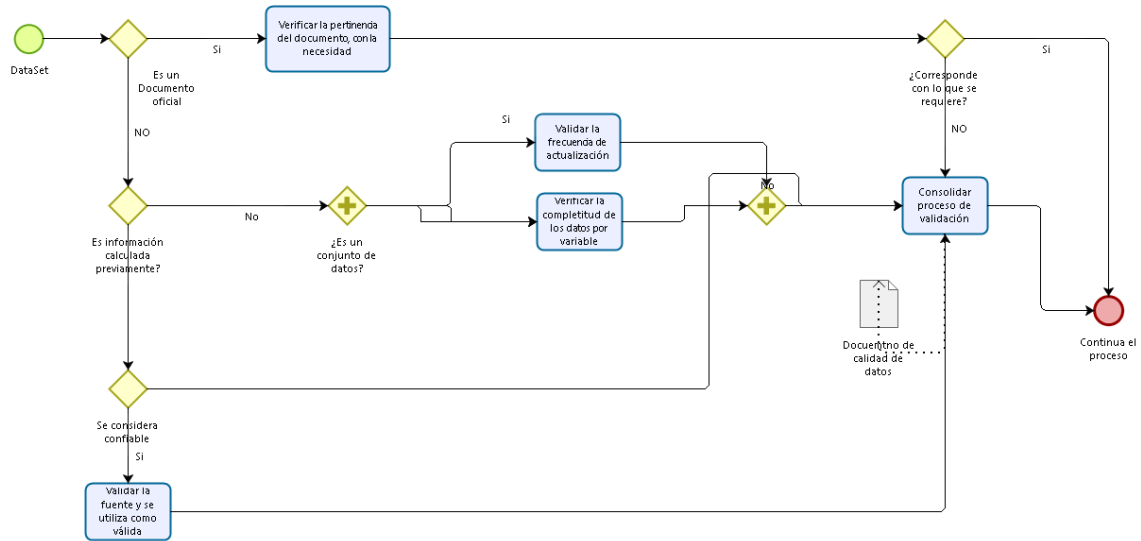


Ilustración 6. Flujo del subproceso de evaluación de calidad



GOBIERNO DE DATOS

Observatorio ODS (objetivos de desarrollo sostenible)

ALIANZA CAOBA

Contenido

Tabla de Ilustraciones.....	139
Control de versiones.....	139
Gobierno de Datos.....	140
Objetivo del documento.....	140
Objetivo de los datos almacenados.....	140
Metas propuestas.....	140
Responsables del gobierno de datos.....	140
Seguridad (Gestión de la Seguridad de Datos).....	141
Estructura de la información (Administración de la arquitectura de Datos).....	142
Datos - Definiciones generales.....	143
Propuesta de comunicación.....	145
Proceso de ajuste de políticas.....	145

Tabla de Ilustraciones

Ilustración 1 Estructura de administración de la política.....	141
Ilustración 2- Esquema de distribución de información.....	143
Ilustración 3- Flujo de ajuste y revisión del gobierno de datos.....	145

Control de versiones

Fecha	Responsable	Detalle del cambio
01-09-2017	Vivian Rivera - Carlos Pulido	Primera versión
20-11-2017	Vivian Rivera - Carlos Pulido	Versión Final

Gobierno de Datos

Este documento lista las directrices, mejores prácticas y recomendaciones de manejo de datos para la información relacionada con el observatorio de datos, su almacenamiento, el procesamiento y la visualización, enfocado en los roles que interactuarán con este. Como parámetro de referencia se utilizará el DAMA (<https://www.dama.org>) y su metodología para el gobierno de datos.

37. Objetivo del documento

Definir los lineamientos y políticas que permitirán estructurar la información que se almacenará en el Observatorio de Datos de los ODS así como mecanismos de control y comunicación que se deber surtir para hacer cambios en la estructura y en los tipos de datos que se almacenarán.

38. Objetivo de los datos almacenados

Ser el insumo para la visualización de gráficas que permitan describir el comportamiento de los datos.

Permitir realizar modelos analíticos que acompañen la visualización de datos para la toma de decisiones.

Soportar las decisiones que se tomen sobre los datos resultados de los análisis realizados.

Poder ser auditados para conocer los momentos en que se han intervenido.

39. Metas propuestas

Tener información con una completitud superior al 80% de los datos en los primeros 6 meses e ir incrementando 1% semestral.

Realizar revisión de fuentes de datos anualmente para actualizar y mejorar la calidad de los datos ya existentes o la inclusión de información nueva.

Garantizar la existencia de la auditoría de los datos de tal forma que se pueda hacer seguimiento del ingreso de los datos y la fecha de actualización de estos.

Garantizar la normalización de los datos que lo requieren para hacerlos comparables en el modelo.

Que cada conjunto de datos incluya una variable de control que permita identificar si la información es pública o no.

40. Responsables del gobierno de datos

Con el propósito que este documento tenga el impacto esperado en el observatorio de datos se debe asignar una figura directiva que tenga el poder dentro de la organización de liderar y hacer cumplir la aplicación de este, este cargo directivo preferiblemente deberá tener dominio en la generalidad de los proyectos que se incluirán en el observatorio y que articule los temas tecnológicos. Por esto se propone un primer nivel al Arquitecto de SW y HW del Centro de Excelencia Caoba, o un cargo del mismo nivel o superior.

Dentro de las funciones del líder de gobierno de datos está:

- Identificar y expresar, en conjunto con los líderes de los proyectos, las necesidades de datos que cada proyecto tiene o tendrá, para verificar si es aplicable el gobierno de datos y validar los ajuste a los que haya lugar y proceder de conformidad.
- Solicitar a cada líder de proyecto el cumplimiento de la política y realizar revisiones periódicas aleatorias y de control sobre los datos.
- Citar al comité de datos según identifique la necesidad de hacerlo.

- Actualizar la política de datos según las solicitudes del comité.

Para apoyar la administración del líder de gobierno de datos, se propone que se defina un comité de datos, que se encargará de:

- Autorizar la modificación de las directivas de la política de datos plasmada en este documento.
- Direccional la política del gobierno de datos, definiendo las prioridades que se deben tener y los datos que se deben depurar o actualizar,
- Incluir nuevas directrices o modificar las existentes.

A continuación, un esquema propuesto de la interacción entre los niveles de control del gobierno de datos, el observatorio y los proyectos que se articularán en el observatorio.

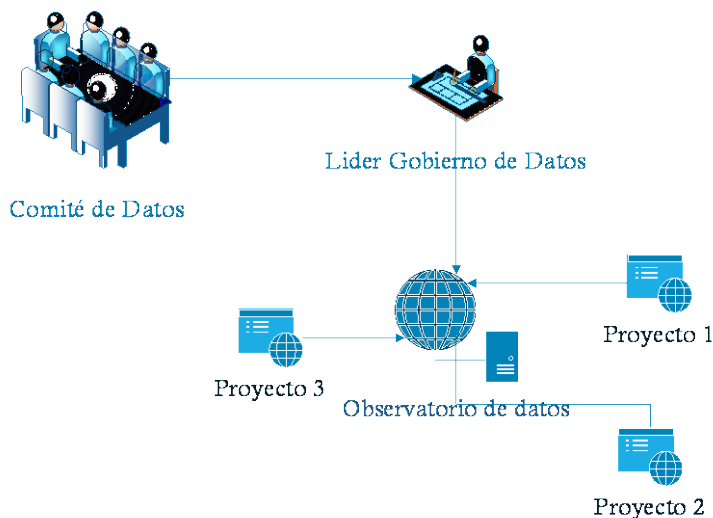


Ilustración 7 Estructura de administración de la política

41. Seguridad (Gestión de la Seguridad de Datos)

El esquema propuesto se basa en los diferentes roles, que permitirán una distribución de las labores y una organización para realizar los procesos de la mejor manera. Adicionalmente se enfocará en el manejo que se hace de usuarios en la herramienta seleccionada para la prueba prototipo.

El esquema de seguridad se plantea en 2 niveles, uno a nivel de roles por usuarios y otro temático, en este caso llamaremos a los grupos temáticos organizaciones que tienen asociados grupos de usuarios y estos grupos unos roles sobre la organización, este esquema nos permitirá suficiente flexibilidad según las necesidades.

A continuación, un listado de usuario y roles propuestas:

Rol	Descripción	Permisos sobre
-----	-------------	----------------

		Organizaciones	Grupos	Conjunto de datos	Consulta
Sysadmin	Acceso completo a todos los niveles de la información. Creación y eliminación	X	X	X	X
Mantenimiento Junior	Tendrá privilegios sobre los grupos y los conjuntos de datos de las organizaciones que el Sysadmin defina.		X	X	X
Visualizador	Usuario que acceden con acceso de consulta a los gráficos, no requiere registro alguno o permiso específico.			X	X

41.1. Estructura de la información (Administración de la arquitectura de Datos)

A diferencia de una bodega de datos tradicional en donde se hacen agrupaciones según el nivel de granularidad definido por la gerencia o por las personas que la utilizarán, esta bodega pretende almacenar información para realizar modelos analíticos, por lo que deben poderse consultar con el mayor nivel de detalle permitiendo hacer un análisis mejor enfocado, por esto la estructura que se propone permite la convivencia de dos modelos, el primero para los modelos analíticos y el segundo para la generación de gráficos estadísticos.

La estructura propuesta de los datos se sugiere teniendo en cuenta el modelo de funcionamiento de la Alianza CAOBA, puesto que está enfocada a proyectos y que pueden ser de mediano o largo plazo; para esto se propone un eje central de información que puede ser utilizada para cualquier proyecto, y sobre los que girarán los Objetivos de Desarrollo Sostenible, estos datos serán:

1. Los de uso general, información que puede ser utilizada como insumo por varios objetivos.
2. Las de uso específico, aquella que tiene información tan puntual que no puede ser

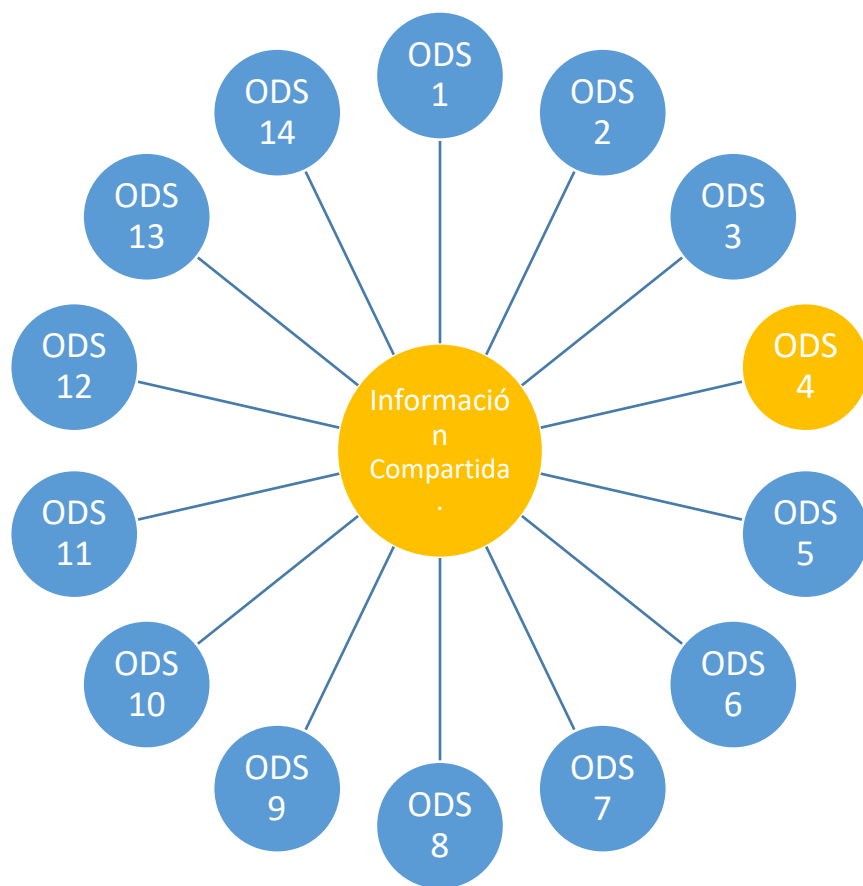


Ilustración 8- Esquema de distribución de información

41.2. Datos - Definiciones generales

Se proponen los siguientes mecanismos de homogenización de datos que aplicará principalmente la para la información compartida:

1. Los datos deben estar anonimizados.
2. Las definiciones de códigos y siglas se deben estandarizar según las necesidades de almacenamiento tratando de optimizar el almacenamiento, claridad del valor almacenado y la posibilidad de que tome todos los valores deseados. Estas definiciones se deberán realizar enfocadas a la estandarización de las fuentes y según las prioridades que se manejen, por ejemplo adoptar la definición que el gobierno hace en sus definiciones como el primer nivel, y luego se pueden incluir valores adicionales que no están incluidos en el estándar nacional o definiciones internas de Caoba.

Ejemplo, Género:

Actualmente el detalle de la información asociada a esto incluye la preferencia sexual por esto se puede proponer un almacenamiento que permita almacenar valores diferentes a Masculino y Femenino, por ejemplo "I" de indefinido o "A" de ausente.

Cada codificación definida debe hacer parte del diccionario de datos y debe ser revisado y actualizado cada vez que se realice un proceso de carga con información nueva.

Datos - Codificación de variables

Se define el siguiente mecanismo para la definición de nombres para facilitar la identificación de la información que se almacena en esta.

- Utilización del "_", este será el único separador permitido para los nombres de variables, para identificar los cambios de palabra.
- Los nombres que se definan deben contener por lo menos las 3 primeras letras de cada palabra que conformen el título de la variable que describe la información que se almacenará, garantizando que sea comprensible su interpretación.
- Los nombres de variables en su versión corta no deben superar los 30 caracteres.
- Se debe definir un diccionario de variables, que permita identificar a que se refiere cada variable.

Datos – Almacenamiento por tipo de dato

Se listará los tipos de datos de datos más comunes y la forma óptima de su almacenamiento, basado en las siguientes reglas generales:

- Siempre se debe preferir tener el mínimo de información en un mismo campo.
- Las longitudes de los campos deberán ser lo más ajustada posible más un margen de crecimiento estimado.

Cadenas de caracteres

- nombres propios: Natalia Rivera siempre preferiremos almacenar en una columna el nombre: Natalia y en otra el apellido: Rivera, si se tiene la posibilidad de tener mas de un nombre cada uno se almacenará en una columna diferente al igual que cada apellido.
- Las longitudes de los campos serán del más largo y se le debe adicionar 20 caracteres para tener margen de acción con mínimo impacto.
- Todas las cadenas de texto se deben almacenar en mayúsculas o en minúsculas, pero no combinadas y todas las cadenas se deben comportar igual.

Fechas

- Los formatos de fecha se deben almacenar en 3 campos diferentes, uno para el día, otro para el mes y otro para el año.

Números

Los números se clasificarán en dos: enteros y con decimales.

- Se debe seleccionar el tipo de dato según el valor que se almacenará con el fin de no perder información
- La precisión se debe asignar según los datos existentes y se debe incluir un margen de crecimiento basado en las estadísticas que describen el comportamiento.

Datos – Metadatos

- Se debe identificar que variables contienen información que puede ser divulgada según la fuente que se obtuvo, esto se incluye los datos anonimizados.
- Tipo de información, los descriptores de variables deben incluir un indicador de información sensible o no.
- Fechas de inserción, se deben incluir descriptores que permitan identificar en qué momento se ingresó la información al repositorio y la fecha de generación de los datos.

- Fechas de actualización, es necesario tener un registro de actualización de los datos indicando la fecha en la cual se actualizó este dato.

42. Propuesta de comunicación

Para dar a conocer las políticas definidas, se tomarán las siguientes acciones:

- Incluir una descripción de los datos de tal forma que se pueda identificar los valores que puede tomar la variable y lo que representa cada valor.
- Cada nueva política métrica o definición se comunicará por correo electrónico a los integrantes del comité de datos.
- Se debe tener actualizado una wiki, que será el directorio de datos que defina y documente los lineamientos definidos y los cambios realizados.

43. Proceso de ajuste de políticas

Cada usuario que interactúe con el observatorio al nivel de conjuntos de datos podrá realizar propuestas de ajustes en el gobierno de datos y este deberá pasar por un proceso de revisión, aprobación e implementación así:

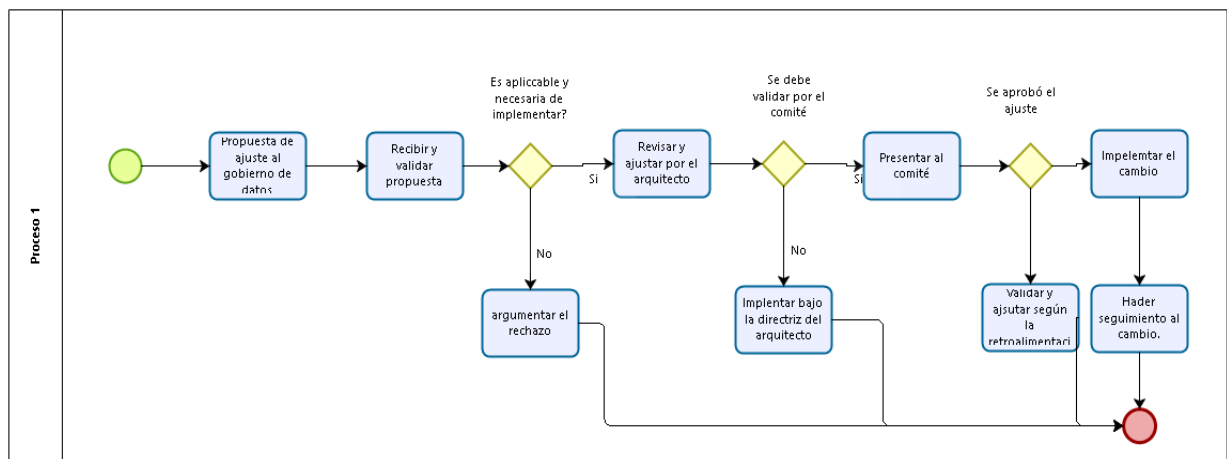


Ilustración 9- Flujo de ajuste y revisión del gobierno de datos

Cada año se debe hacer una revisión del gobierno de datos para ajustarlo según la evolución y la práctica haya identificado, en esta revisión se pueden proponer ajustes a todo nivel del documento que pueden implicar eliminar actividades, incluir nuevas o ajustes el orden de ejecución. Cada una de estas sugerencias deberá cubrir el flujo previo mencionado. Teniendo en cuenta este proceso continuo de revisión se incluye como anexo el proceso de implementación de la prueba piloto en la que se aplicará todo este gobierno de datos desde el proceso de instalación hasta la finalización de la prueba piloto.

ANEXO 3. Resumen de variables

CAMPO	Descripción	Tipo de dato	Valores faltantes	Porcentaje faltantes	2000-1	2000-2	2001-1	2001-2	2002-1	2002-2	2003-1	2003-2	2004-1	2004-2	2005-1	2005-2	2006-1	2006-2	2007-1	2007-2	2008-1	2008-2	2009-1	2009-2	2010-1	2010-2	2011-1	2011-2	2012-1	2012-2	2013-1	2013-2	2014-1	2014-2	2015-1	2015-2	2016-1	2016-2		
PERIODO	Abierta (ejemplo 20101-20102)	Numérico	0	0,00%	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
ESTU_CONSECUTIVO	Id público del estudiante	Cadena	4	0,00%	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
ESTU_ESTUDIANTE	Indica si realizó la inscripción por medio de un colegio (estudiante) o fue individual	Cadena	0	0,00%																																X	X	X	X	
ESTU_EDAD	Edad del estudiante	Numérico																					X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
ESTU_TIPO_DOCUMENTO	Tipo de Documento del estudiante	Cadena	832.261	10,45%	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
ESTU_GENERO	Género del estudiante	Cadena	22.922	0,25%	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
ESTU_FECHA_NACIMIENTO	Fecha de nacimiento del estudiante	Fecha																																						
ESTU_NACIMIENTO_MES	Mes de nacimiento del estudiante	Numérico	15.554	0,62%																			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
ESTU_NACIMIENTO_ANNO	Año de nacimiento del estudiante	Numérico	15.554	0,62%																			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
ESTU_TIENEETNIA	¿Pertenece usted a un grupo étnico minoritario?	Numérico	12993	100,00%																																				
ESTU_ETNIA	Etnia a la cual pertenece el estudiante en caso de que tenga	Numérico	0	0,00%	X										X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X			
ESTU_LIMITA_MOTRIZ	¿El estudiante presenta discapacidad motriz?	Numérico	1.361.495	100,00%																																X	X	X	X	
ESTU_LIMITA_INVIDENTE	¿El estudiante presenta discapacidad invidente?	Numérico	1.361.495	100,00%																																X	X	X	X	
ESTU_LIMITA_CONDICIONESPECIAL	¿El estudiante presenta discapacidad de condición especial?	Numérico	1.361.495	100,00%																																X	X	X	X	
ESTU_LIMITA_SORDO	¿El estudiante presenta discapacidad de sordera?	Numérico	1.361.495	100,00%																																X	X	X	X	
ESTU_LIMITA_SDOWN	¿El estudiante presenta discapacidad del síndrome de down?	Numérico	1.361.495	100,00%																																	X	X	X	X
ESTU_LIMITA_AUTISMO	¿El estudiante presenta discapacidad de autismo?	Numérico	1.361.495	100,00%																																	X	X	X	X
ESTU_LIMITA_BAJAVISION	¿El estudiante presenta discapacidad de baja visión?	Numérico	1.361.495	100,00%																																	X	X	X	X
ESTU_LIMITA_COGNITIVA	¿El estudiante presenta discapacidad cognitiva?	Numérico	1.361.495	100,00%																																	X	X	X	X
ESTU_LIMITA_SORDOINTERPRETE	¿El estudiante presenta discapacidad sordointerprete?	Numérico	1.361.495	100,00%																																X	X	X	X	
ESTU_LIMITA_SORDONONINTERPRETE	¿El estudiante presenta discapacidad sordo no interprete?	Numérico	1.361.495	100,00%																																X	X	X	X	
ESTU_LIMITA_SORDOCEGUERA	¿El estudiante presenta discapacidad sordoy ceguera?	Numérico	1.361.495	100,00%																																X	X	X	X	
ESTU_VECES_ESTADO_ESTUDIANTIL	Número de veces que ha presentado el examen de estado de la población bachiller	Numérico	108.525	15,90%																																X	X			

