MAPA DE LA SITUACIÓN ACADÉMICA COLOMBIANA A TRAVÉS DEL ANÁLISIS DE LAS BASES DE DATOS DEL ICFES

ANDREA PAOLA SÁNCHEZ PÉREZ

UNIVERSIDAD AUTÓNOMA DE BUCARAMANGA
FACULTAD DE INGENIERÍA
MAESTRÍA EN GESTIÓN, APLICACIÓN Y DESARROLLO DE SOFTWARE
BUCARAMANGA
2020

MAPA DE LA SITUACIÓN ACADÉMICA COLOMBIANA A TRAVÉS DEL ANÁLISIS DE LAS BASES DE DATOS DEL ICFES

ANDREA PAOLA SÁNCHEZ PÉREZ

Trabajo de grado para optar al título de
Magister en Gestión, Aplicación y Desarrollo de Software

DIRECTORA
MARITZA LILIANA CALDERON BENAVIDES
Doctora en Informática y Comunicación Digital

UNIVERSIDAD AUTÓNOMA DE BUCARAMANGA
FACULTAD DE INGENIERÍA
MAESTRÍA EN GESTIÓN, APLICACIÓN Y DESARROLLO DE SOFTWARE
BUCARAMANGA
2020

Nota de Aceptación

_____

_____

_____

_____

Presidente del Jurado

_____

Jurado

_____

Jurado

Bucaramanga, 06 de noviembre de 2020

A Dios, guía y protector.

A Nati y Sebas, tesoros que iluminan mi vida.

A Javi, compañero de vida.

A mi familia, amor y apoyo incondicional.

AGRADECIMIENTOS

A la doctora Liliana Calderón por sus aportes y orientación en la dirección del proyecto.

A los docentes de la maestría por compartir valiosos conocimientos y experiencias.

A la ingeniera Claudia Cáceres por su colaboración en la finalización del proyecto.

CONTENIDO

LISTA DE TABLAS

LISTA DE FIGURAS

LISTA DE ANEXOS

GLOSARIO

ARCHIVO CSV: (CSV: *Comma Separated Values*), tipo de documento en formato abierto que permite el almacenamiento de grandes cantidades de datos estructurados, representados en forma de tabla, en el cual las columnas se separan por comas y las filas por saltos de línea.

CLÚSTER: grupo de elementos con similitud entre ellos.

CRISP-DM: (*Cross Industry Standard Process for Data Mining*), metodología para el análisis de proyectos de minería de datos.

ICFES: Instituto Colombiano para la Evaluación de la Educación.

KDD: *Knowledge Discovery in Databases*, proceso que busca descubrir conocimiento a partir de un repositorio de información.

KNIME: (*Konstanz Information Miner*), plataforma de minería de datos que permite tratamiento de datos y desarrollo de modelos en un entorno visual.

Medida de dispersión: medida estadística que evalúa en qué medida los datos difieren entre sí. Ejemplo: desviación estándar, varianza.

Medida de tendencia central: medida estadística que permite identificar un centro en torno al cual se encuentra ubicado el conjunto de los datos. Ejemplo: media, mediana y moda.

MEN: Ministerio de Educación Nacional

NSE: nivel socioeconómico del evaluado.

Saber 11: prueba estandarizada aplicada a estudiantes de último año de educación media, requisito para obtener el título de bachiller.

Tableau: herramienta software que permite crear visualizaciones de datos *online* de forma gráfica e interactiva.

RESUMEN

Este trabajo presenta el resultado de un estudio de la situación académica colombiana basado en el análisis de los resultados de las pruebas estandarizadas SABER 11. A partir de la revisión de la información obtenida de las bases de datos del ICFES, donde se encuentran publicados los resultados históricos de las pruebas que se han aplicado a los estudiantes colombianos de último grado de formación media en el período 2015-2019; se realiza un análisis descriptivo de los resultados por medio de la aplicación de técnicas de minería de datos y de la metodología CRISP-DM (*Cross Industry Standard Process for Data Mining*), realizando la preparación, comprensión, selección y modelado de datos; permitiendo de esta manera, identificar tendencias, variables personales, académicas, socioeconómicas y familiares que influyen directamente en el desempeño académico de los estudiantes; adicionalmente, con el modelamiento realizado se identifican cuatro clústeres o grupos basados en características similares encontradas.

Se desarrolla una herramienta de visualización que incluye mapas interactivos de Colombia, que presentan la información analizada de forma dinámica y permiten la comprensión geográfica de la misma a nivel regional, departamental y municipal; tanto del total de la población de estudio como de la población de evaluados perteneciente a grupos étnicos.

PALABRAS CLAVE: SITUACIÓN ACADÉMICA COLOMBIANA, BASE DE DATOS ICFES, SABER 11, MAPA INTERACTIVO, MINERÍA DE DATOS.

ABSTRACT

This work presents the result of a study of the Colombian academic situation based on the analysis of the results of the SABER 11 standardized tests. Beginning with a review of the information obtained from ICFES databases, where test historical results applied to Colombian students of last year of high school between 2015 and 2019 are published; a descriptive analysis was carried out through the application of data mining techniques and the methodology for data analysis: CRISP-DM (Cross Industry Standard Process for Data Mining), which includes data preparation, data understanding, data selection and data modeling; allowing this way to identify trends in results; personal, academic, socioeconomic and family variables that influence the academic performance of students; in addition, with the perform of data modeling a cluster analysis was carried out based on attributes similarities which permitted to identify four different clusters that characterize the study population.

With this work, it was also developed a visualization tool that includes Colombian interactive maps and statistics that present the analyzed information in a dynamic way and allow the user to get a geographic understanding of the results at national, regional and city level for both the general population and population belonging to ethnic groups.

KEYWORDS: COLOMBIAN ACADEMIC SITUATION, ICFES DATABASES, SABER 11, INTERACTIVE MAP, DATA MINING.

INTRODUCCIÓN

Conocer la situación académica actualizada de un país es esencial para el mejoramiento de la calidad educativa, dado que proporciona información que permite determinar fortalezas y oportunidades de mejora que orienten el diseño de políticas y direccionen la definición de programas y planes de mejoramiento. Una forma de conocer el estado de la educación es a través del análisis de resultados de pruebas estandarizadas aplicadas a la población estudiantil. El Instituto Colombiano para el Fomento de la Educación Superior (ICFES) es la entidad encargada de evaluar la educación en todos sus niveles.

A nivel nacional, se han desarrollado diversos estudios para determinar la situación académica del país basados en resultados de pruebas estandarizadas; sin embargo, no se cuenta con un análisis actualizado que incluya información consolidada a nivel nacional, regional y departamental que permita identificar el estado de la educación en Colombia. Con el desarrollo de este trabajo, se aplican técnicas de minería de datos para la realización un análisis descriptivo de la situación académica colombiana a partir de los resultados obtenidos en las pruebas Saber 11 aplicadas por el ICFES, presentadas por los estudiantes de último nivel de educación media, entre los años 2015 y 2019 y se implementa una herramienta de visualización que contiene la información clasificada y organizada permitiendo realizar su análisis desde una perspectiva geográfica a nivel regional y departamental, identificando.

El documento se encuentra organizado de la siguiente manera. En el capítulo 1, se presenta la formulación de la investigación que incluye el planteamiento del problema, la justificación de la realización del trabajo y los objetivos propuestos en este proyecto. El capítulo 2 incluye el marco teórico y el estado del arte en cuanto al análisis de la situación académica en Colombia y en el mundo. El capítulo 3 presenta la metodología propuesta para el desarrollo del trabajo. El capítulo 4 detalla la aplicación de la metodología y sus fases, así como los resultados obtenidos y la herramienta desarrollada.

1. FORMULACIÓN DE LA INVESTIGACIÓN

En este capítulo se realiza una descripción del problema encontrado de la falta de conocimiento de la situación académica actual colombiana y se plantea su solución, identificando los objetivos que se cumplirán con el desarrollo del proyecto.

1.1. PLANTEAMIENTO DEL PROBLEMA

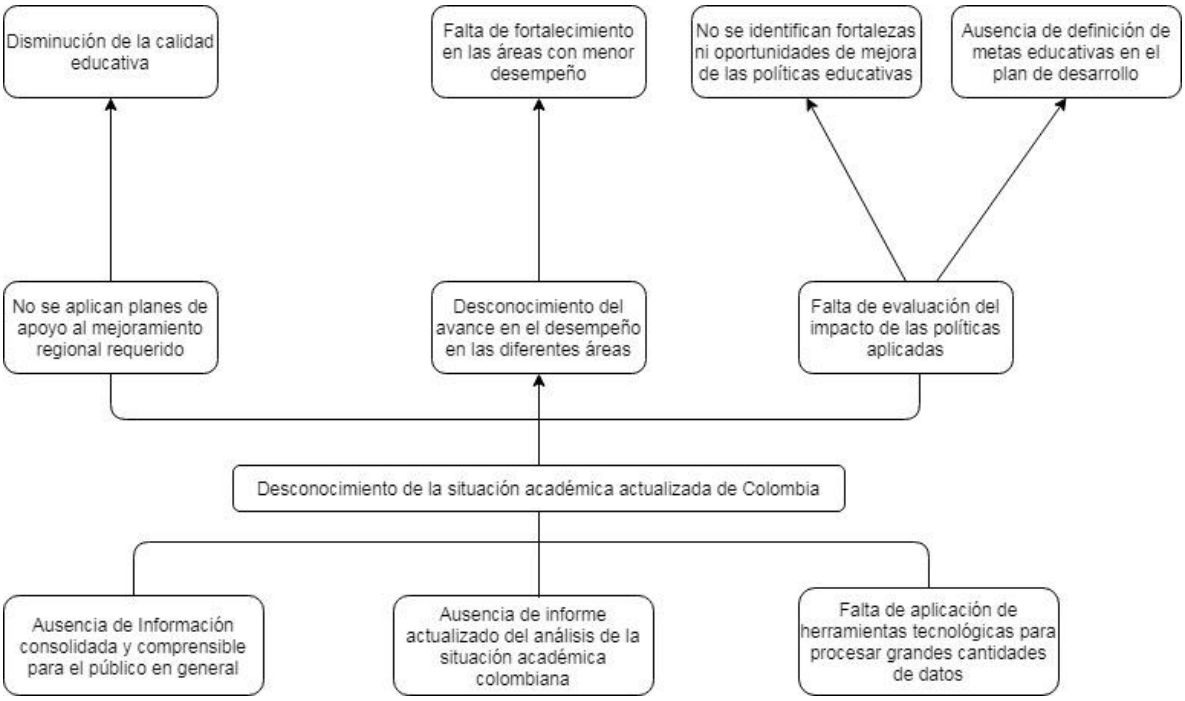
Conocer el estado de la situación académica actualizada de un país o región, es una herramienta esencial para el mejoramiento de la calidad educativa, ya que proporciona información que permite determinar fortalezas y oportunidades de mejora que orientan el diseño de políticas y la definición de programas por parte de los organismos rectores del sector, así como también, la elaboración de planes de mejoramiento por parte de las mismas instituciones escolares. Además, es fuente importante para la realización de investigaciones educativas e innovaciones pedagógicas. (Secretaría de Educación de Bogotá, 2004)

Una de las formas de evaluar la situación académica de un país o región es la evaluación de los resultados de la aplicación pruebas estandarizadas a estudiantes. En Colombia, la institución encargada de realizar la evaluación a la educación es el ICFES, que cada año aplica estas evaluaciones a miles de estudiantes del país, y genera grandes volúmenes de datos con información socioeconómica y puntajes obtenidos por los estudiantes.

En Colombia se han desarrollado diversos estudios para analizar la situación académica del país a partir de los resultados de pruebas aplicadas por el ICFES (Rodríguez, Maldonado, & Barrera, 2012). De hecho, el ICFES ha publicado reportes estadísticos de los resultados de los exámenes que aplica tanto a nivel de instituciones educativas como a nivel de entidades territoriales (ICFES, 2020b) y se han desarrollado diversos estudios para analizar el impacto de ciertas variables en el rendimiento de los estudiantes en las pruebas. Sin embargo, no se cuenta con un análisis actualizado que incluya información consolidada a nivel nacional que

permita determinar la situación académica colombiana. La figura 1 presenta el árbol causa-efecto que genera esta situación.

Figura 1. Árbol de problemas: causas-efectos



Fuente: Elaboración propia

En Colombia cada año, más de 500.000 estudiantes en su último año de educación media presentan pruebas y sus resultados contienen datos de puntajes e información socioeconómica que puede ser analizada y servir como fuente de información para la aplicación de planes de mejoramiento.

En Colombia, no se cuenta con una herramienta que suministre información actual, organizada y clasificada de la situación académica de Colombia, que provea información que apoye la toma de decisiones en los programas de fortalecimiento en las áreas de menor desempeño y construcción de indicadores de calidad de la educación que permitan ajustar las políticas de calidad, tanto a nivel nacional como regional.

1.2. JUSTIFICACIÓN

> Los datos son un insumo para la actividad investigativa, académica y productiva, para la toma de decisiones públicas y para la garantía de derechos. Son distintos a otros activos económicos en la medida en que su valor se materializa con el procesamiento y obtención de la información y el conocimiento intrínseco que se obtiene al combinarlos entre sí. (European Data Protection Supervisor, 2014).

A partir del estudio de las bases de datos de resultados de exámenes estandarizados publicados en el FTP del ICFES y en base en el análisis de datos y su estudio desde una perspectiva geográfica se pueden identificar hechos, patrones ocultos de comportamiento de variables, correlaciones y/o tendencias que servirán como instrumento de análisis de la situación educativa en el marco del desarrollo de competencias básicas; permitiendo identificar avances o retrocesos en las áreas, posibilitando de esta forma la realización de comparativos de desempeño tanto a nivel nacional como a nivel regional.

El desarrollo de este proyecto aporta información útil a las secretarías de educación departamentales y/o municipales e instituciones educativas en el área de fortalecimiento pedagógico, ya que permitirá evaluar el impacto de los procesos pedagógicos, aportando información que apoye la toma decisiones en cuanto a los planes de formación, acompañamiento pedagógico y metas a trazar desde el plan de desarrollo municipal y regional. De igual forma, el disponer de esta información organizada permitirá a diferentes ciudadanos investigadores desarrollar propuestas para el mejoramiento de la calidad académica basados en los resultados de los análisis de tendencias y conocimiento descubierto.

Adicionalmente proporcionará información sobre los factores que inciden en el rendimiento académico, identificando fortalezas y oportunidades de mejora de los procesos de aprendizaje; lo cual aportará conocimiento útil que podrá ser utilizado para el diseño de programas y políticas de mejoramiento de la educación a nivel regional y nacional.

1.3. PREGUNTA DE INVESTIGACIÓN

La pregunta de investigación es: ¿Cómo medir el estado actual de la situación académica en Colombia a partir del estudio de las bases de datos de ICFES?

1.4. OBJETIVOS

1.4.1. Objetivo general

Desarrollar un mapa interactivo *online* de la situación Académica de Colombia a partir del análisis de los resultados de las pruebas SABER 11 entre los años 2015 y 2019, con el fin de determinar la situación académica colombiana a nivel nacional, regional y departamental.

1.4.2. Objetivos específicos

Para el cumplimiento del objetivo general se plantean los siguientes objetivos específicos:

- Realizar un estudio de los resultados de exámenes académicos estandarizados a través de la consulta del repositorio de las bases de datos del ICFES con el fin de identificar los diferentes datos que se van a incluir en el análisis de información.
- Realizar un análisis descriptivo de los datos basado en la aplicación de la metodología para minería de datos CRISP-DM que permita identificar características comunes de los datos y clasificarlos en grupos.
- Diseñar, desarrollar la herramienta de visualización que incluya el mapa interactivo de la situación académica colombiana basado en el análisis descriptivo y en los datos analizados.
- Realizar un análisis de la situación académica de Colombia basado en la información analizada, identificando tendencias, situación académica departamental y/o regional, entre otros.

2. REVISIÓN BIBLIOGRÁFICA Y MARCO TEÓRICO

En este capítulo se presentan algunos trabajos que se han realizado para conocer la situación académica de un país; adicionalmente, se presenta la contextualización de la forma en que se abordará la problemática, donde se describen las pruebas estandarizadas realizadas por el ICFES, el concepto de minería de datos, la metodología de desarrollo, así como un marco tecnológico que enuncia generalidades de las herramientas a utilizar en el desarrollo del proyecto.

2.1. ESTADO DEL ARTE

A continuación, se presentan los diferentes avances realizados en el análisis de la situación académica elaborados a partir de resultados de las pruebas estandarizadas educativas aplicadas a estudiantes tanto a nivel nacional como internacional.

2.1.1. Análisis de la situación académica en el mundo

A nivel internacional, se han desarrollado análisis y se han implementado aplicaciones web que permiten visualizar la evolución de los resultados de las pruebas *PISA* (pruebas que evalúan las áreas de lectura, matemática y ciencias naturales y se aplican a estudiantes de 15 años en diferentes países pertenecientes a la OECD). Una de estas aplicaciones es *Our world in data* (Roser, Nagdy, & Ortiz), sitio web que permite visualizar a través de mapas políticos, los resultados en las diferentes áreas, que han tenido todos los países que han participado a través del tiempo, permitiendo acceder a los datos fuentes de las gráficas. Esta aplicación permite que cualquier usuario con acceso a internet pueda visualizar gráficamente los resultados de las pruebas, identificar dónde están los mejores estudiantes y contar con una herramienta que permita realizar un análisis de la evolución de los resultados a través del tiempo.

En 2016 Koskela (Koskela, 2016) realizó un análisis del desempeño de los estudiantes de Finlandia entre los años 2003 y 2012 en las pruebas PISA para tratar de comprender el descenso en los resultados obtenidos en 2012. Usando minería de datos, específicamente algoritmos no supervisados, identificó *clústeres* para los años en estudio y se identificaron las características de los grupos que mejor y peor desempeño tuvieron.

En 2018 Chi Poot desarrolló ALUSE, una plataforma digital que permite determinar las variables que tienen mayor relevancia en las pruebas académicas *PLANEA*[1] y *PISA*. "La herramienta determina las variables que tienen mayor relevancia para los resultados académicos de los alumnos; los datos para el análisis se obtuvieron de datos abiertos de una muestra representativa de *PLANEA* 2015 a nivel de secundaria con aproximadamente 144000 registros" (Chávez, 2019). Su objetivo fue el de ofrecer información contextualizada para la mejora de los procesos de enseñanza, comunicar a la sociedad sobre el estado de la educación y aportar a las autoridades educativas asesoría relevante y útil para la planeación, programación y operación del sistema educativo mexicano.

2.1.2. Análisis de la situación académica en Colombia

Para la realización del análisis de la calidad académica, departamentos y municipios, con el apoyo de sus secretarías de educación, han desarrollado análisis y estudios particulares que sirvan de apoyo en la toma de decisiones; en otras regiones el ICFES realiza el análisis focalizado y lo socializa en la región de interés; de forma complementaria se han desarrollado investigaciones independientes que han analizado la situación académica a partir de resultado de exámenes estandarizados, como se presenta a continuación:

En 2015 Martínez & Turriago (Martínez & Turriago, 2015), desarrollaron un análisis de distribución geográfica y espacial de los resultados de las pruebas Saber 11 del

[1] PLANEA: Plan Nacional para la Evaluación de los Aprendizajes. Prueba diseñada para aportar información relevante del logro educativo, permite conocer la medida en que los estudiantes alcanzan aprendizajes clave al término de la Educación Secundaria en México.

ICFES entre los años 2005 y 2012 a nivel nacional; basados en información publicada por el ICFES y el DANE se analizaron los puntajes a nivel municipal y se contrastaron con variables específicas (relación alumnos/docentes, necesidades básicas insatisfechas, etc.), para determinar su influencia en la calidad de la educación. A partir del análisis estadístico se identificó la relación directa entre los niveles de pobreza y los resultados deficientes en las pruebas de estado al igual que a mayor cantidad de docentes por estudiante mejor rendimiento en las pruebas. De forma complementaria, se identificaron los departamentos con mejor desempeño. En este estudio se presentó la evolución de los resultados de las pruebas ICFES por municipios entre los años 2005 y 2012, representada en mapas elaborados en el software de análisis geoestadístico Geoda.

En 2015 Blanco (Blanco V. , 2015a) realizó un análisis del desempeño académico de los estudiantes del departamento de Cesar que presentaron el examen de estado Saber 11 en el año 2012; este análisis se desarrolló a través de un estudio de minería de datos en educación basado en la información publicada por el ICFES siguiendo la metodología *CRISP-DM* y la aplicación de un algoritmo de agrupamiento; el desarrollo de esta investigación permitió identificar y caracterizar los grupos de estudiantes que obtuvieron los diferentes niveles de desempeño en la prueba; de forma complementaria, se creó un mapa electrónico navegable en la herramienta *OpenStreetMap* donde se visualizan los resultados obtenidos por municipio (Blanco V. , 2015b).

En 2015 el Instituto Colombiano para la Evaluación de la Educación (ICFES, 2015a) desarrolló un informe nacional de resultados de Colombia en las pruebas *PISA*. Las pruebas *PISA* (programa para la evaluación internacional de alumnos) son unas pruebas estandarizadas que evalúan la calidad de la educación en países miembros de la Organización para la cooperación y el desarrollo económico (OCDE). En este informe se presentaron los resultados de Colombia en comparación con las demás economías participantes teniendo en cuenta los resultados obtenidos desde el año 2006 hasta 2015 con el fin de “contribuir a las discusiones académicas y de política educativa, y promover los esfuerzos para consolidar una formación de alto nivel para las generaciones presentes y futuras” (ICFES, 2015a).

En el 2015 el ICFES inició la divulgación regional de los resultados de las pruebas SABER a través de encuentros los regionales en donde se presentan los puntajes generales y niveles de desempeño de la región comparados con los resultados promedio del país. (ICFES, 2015b)

En 2016 la secretaría de educación municipal de Cali (Secretaría de Educación Municipal de Santiago de Cali, 2016) tomando como fuente de información las bases de datos del ICFES, realizó un análisis comparativo - descriptivo de los resultados de las pruebas SABER entre los años 2012 y 2015 en las cuatro áreas de desempeño. La finalidad de este proyecto fue la de realizar el diagnóstico del estado actual del desarrollo de competencias y visualizar el impacto de los procesos pedagógicos que se llevaron a cabo y proporcionar elementos para la toma de decisiones a nivel institucional, establecer fortalezas y debilidades en el desarrollo de competencias para aplicar planes de mejoramiento.

En 2017 el ICFES publicó en su sitio web un *reporteador* de históricos (ICFES, 2017b), que según su definición permitía visualizar los resultados históricos por colegios, municipios, departamentos para las pruebas Saber 11 aplicadas entre los años 2000-2014 y las pruebas Saber Pro aplicadas antes del 2011. Sin embargo, esta aplicación, no se encuentra actualmente en funcionamiento, aunque se puede cargar, no se ejecuta en la web.

De forma complementaria, en el 2018 el ICFES publicó en su sitio web una aplicación (ICFES, 2018a) que permite la visualización de los resultados de las pruebas Saber 3°,5° y 9° del año 2016 y la generación automática de reportes comparativos de los resultados obtenidos por una ciudad vs. los resultados obtenidos en todo el país teniendo en cuenta los puntajes obtenidos y la situación socioeconómica de los estudiantes.

En 2018 Jiménez (Jiménez, 2018) , desarrolló un análisis del desempeño de las pruebas saber-pro en estudiantes de ingeniería en Antioquia, en donde se aplicó la minería de datos para encontrar patrones que se relacionan con el puntaje en los diversos módulos evaluados en la prueba mencionada; con el fin de evaluar la

incidencia del entorno económico, demográfico y social de los estudiantes en su rendimiento académico.

En 2019 la secretaría de educación de Boyacá (Secretaría de Educación de Boyacá, 2019), desarrolló un histórico de las pruebas SABER entre los años 2012-2018 tomando como fuente de información las bases de datos del ICFES. El propósito fundamental fue el de identificar el estado real del proceso formativo en establecimientos educativos del departamento, identificando las áreas de mejor desempeño y las áreas con dificultades, con el fin de plantear acciones estratégicas a los planes de mejoramiento institucional y de esta forma mejorar la calidad de la educación en el departamento.

A finales del 2019, el ICFES realizó un informe nacional del examen Saber 11 a partir de los resultados obtenidos en 2018 e históricos desde el 2014 (ICFES, 2019); en este informe se analizaron los resultados para el puntaje global y puntajes por prueba; Los resultados del 2018, además de contar con los resultados globales y por prueba del examen, se desagregaron a partir de cuatro características: género, grupos de comparación y sector del establecimiento, zona y sector del establecimiento y por entidad territorial certificada; buscando explicar comportamientos en el desarrollo de las competencias de los estudiantes a través de su formación en la educación media.

Actualmente, el ICFES pone a disposición del público en general diversos reportes de resultados de los exámenes realizados (ICFES, 2020b), en los cuales presenta los resultados históricos clasificados por planteles, por secretarías de educación y por establecimientos educativos. Particularmente, los informes de entidades territoriales incluyen comparaciones a través de las cuales se puede determinar los resultados de la entidad territorial en una aplicación particular del examen con respecto a los resultados del país, establecimientos privados, oficiales rurales y oficiales urbanos de la entidad territorial.

De forma complementaria, el ICFES ha venido realizando convocatorias para impulsar proyectos de investigación relacionados con la calidad de la educación a partir de análisis y estadísticas de los resultados de todas las pruebas que aplica el

ICFES. Algunos de los proyectos de investigación seleccionados han orientado sus estudios al análisis de factores que influyen en el desempeño académico de los estudiantes en las pruebas SABER 11, tales como el estudio realizado por Gaviria y Posada (Gaviria & Posada, 2019) que analizan la influencia de los niveles de contaminación en los resultados, o la investigación realizada por Molina, Morales y Rodríguez (Molina, Morales, & Rodríguez, 2019) que analiza la incidencia de las TIC en el mejoramiento de resultados, entre otros.

2.2. MARCO TEÓRICO

2.2.1. ICFES y evaluación de la educación

> La evaluación permite establecer el nivel de los aprendizajes que se espera que logren los estudiantes mediante su paso por el proceso educativo. Adicionalmente, brinda información a rectores, docentes, padres de familia y estudiantes para que identifiquen las destrezas, habilidades y competencias que los estudiantes desarrollan durante la trayectoria escolar; al detectar fortalezas y debilidades, permite valorar el impacto de los procesos educativos en el desarrollo de competencias de los estudiantes y definir acciones de mejora en todos los niveles. (Departamento Nacional de Planeación, 2019)

El ICFES es la entidad que diseña, desarrolla, aplica y califica las evaluaciones en los niveles educativos de básica primaria, básica secundaria, media vocacional y pregrado; los resultados de estas pruebas determinan el estado de las competencias de los estudiantes, a la vez sirven de herramienta para el mejoramiento de la calidad educativa de las instituciones educativas. A este conjunto de pruebas se les conoce como Pruebas Saber y se aplican de la siguiente forma: para los bachilleres la prueba Saber 11, para los profesionales la prueba Saber PRO, para técnicos y tecnólogos la prueba Saber TyT, y pruebas de calidad durante el proceso de formación de los estudiantes de colegio como son SABER 3, 5 y 9 que permiten identificar las competencias y habilidades que todos los

estudiantes colombianos desarrollan durante su trayectoria escolar. La tabla 1 presenta un resumen de estas pruebas.

Tabla 1. Resumen Pruebas SABER

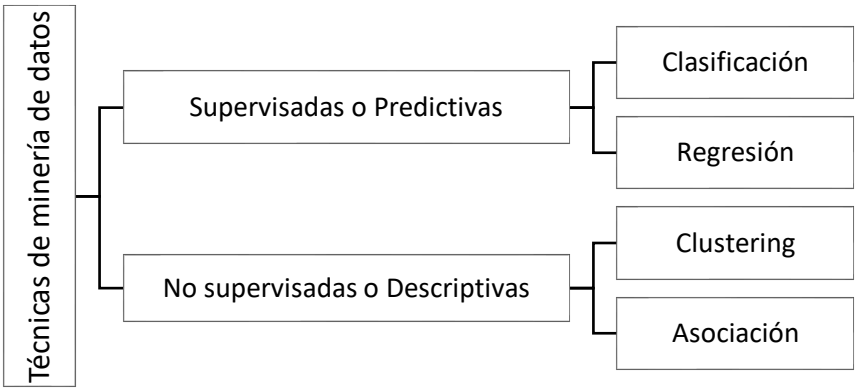
Prueba	**Definición**	**Áreas de conocimiento evaluadas**
Saber 3	Evaluación de estudiantes que se encuentran cursando el grado tercero.	- Lenguaje - Matemáticas
Saber 5	Evaluación de estudiantes que se encuentran cursando el grado quinto.	- Ciencias - Lenguaje - Matemáticas - Competencias ciudadanas
Saber 9	Evaluación de estudiantes que se encuentran cursando el grado noveno.	- Ciencias - Lenguaje - Matemáticas - Competencias ciudadanas
Saber 11	El Examen Saber 11 es presentado por tres tipos de evaluados: - *Estudiantes*: examinandos que presentan el examen a través de una institución educativa y están en el último año de educación media. - *Validantes*: examinandos que presentan el examen para validar su bachillerato. - *Individuales*: examinandos que no presentan a través de un colegio.	- Lectura crítica - Matemáticas - Ciencias Naturales - Sociales y ciudadanas - Inglés
Saber TyT	Evaluación para los programas de formación técnica y tecnológica.	-Competencias genéricas -Competencias específicas
Saber Pro	Evaluación para los programas de formación profesional.	- Competencias genéricas - Competencias específicas

Fuente: Elaboración propia

2.2.2. Minería de datos

La minería de datos es una disciplina que tiene como objetivo el descubrimiento de conocimiento útil a partir de grandes cantidades de datos, por medio de algoritmos, búsqueda e identificación de patrones, tendencias, desviaciones y otros indicadores que extraen conocimiento de grandes repositorios de datos (Blanco V. , 2015b). Existen diferentes técnicas para la realización del análisis de datos. La figura 2 presenta su clasificación.

Figura 2. Técnicas de minería de datos



Fuente: Elaboración propia

- Las técnicas predictivas o supervisadas: se orientan a predecir valores de salida de una variable objetivo. La predicción de la variable objetivo se realiza a partir de un conjunto de variables predictoras, es decir que induce una relación entre un atributo de interés y las otras variables.

  Las técnicas supervisadas son:

  - Clasificación: consiste en examinar las variables predictoras para identificar su clase o categorías las cuales se deben definir previamente. Así se puede catalogar cada nuevo elemento en una clase definida.
  - Regresión: consiste en examinar las variables predictoras para realizar una predicción numérica o regresión.

- Las técnicas descriptivas o no supervisadas: se orientan a acciones descriptivas como descubrir patrones o describir datos; esas técnicas proporcionan información sobre las relaciones entre los datos y sus características, las técnicas no supervisadas son:

    - *Clustering*/Segmentación: Permite segmentar elementos heterogéneos en subgrupos homogéneos, se identifican características para segmentar grupos con gran similitud entre sí y muchas diferencias con los de otros grupos. No requiere de predefinición de clases.
    - Asociación: identifica atributos o elementos que tiene algún valor de asociación, busca co-ocurrencias entre los registros.

    Los métodos más usuales en la aplicación de tareas no supervisadas son:

    - Análisis de correlación: Determina la relación existente entre dos o más variables de estudio, permite obtener como resultado las relaciones entre conceptos de grupos diversos seleccionados. Para su interpretación es necesario contar con el conocimiento acorde al estudio y manejar los conceptos relacionados.
    - *K-means*: Busca similitudes entre los registros por medio de medidas de distancia, de tal manera que los elementos en el mismo clúster están cercanos entre sí.
    - *A priori*: identifica conjuntos de reglas más frecuentes en distintas transacciones u operaciones, funciona a través de la generación de candidatos y reglas de asociación booleanas. Su propiedad principal es que para que un indicador sea frecuente, todos sus subconjuntos no vacíos cuenten con ese indicador frecuente.

La minería de datos ha mostrado sus ventajas en diferentes áreas como negocios, gobierno, medicina, redes sociales, educación, entre otros. Específicamente en la educación, la minería de datos es utilizada para tener una mayor comprensión de los estudiantes, evaluar su progreso y los entornos educativos que aprenden, identificar el éxito o fracaso de estrategias de enseñanza y generar un

discernimiento más profundo del contexto educativo entre otros aspectos (Romero & Ventura, 2010).

El estudio del rendimiento puede darse a través de la predicción y de la descripción, como se mencionó anteriormente, el objetivo de la predicción es estimar el valor desconocido de una variable que describe al estudiante (Rodríguez, Maldonado, & Barrera, 2012) y el objetivo de la descripción es proporcionar la caracterización de un conjunto de datos en los cuales es posible encontrar subgrupos que permitan comprender y explicar el comportamiento del total de los datos.

Este conocimiento descubierto a partir de los datos se denomina *KDD* (*Knowledge Data Discovery*), el cual es un proceso de varias fases que incluye: entendimiento del negocio, preparación de los datos, modelado, evaluación e implementación.

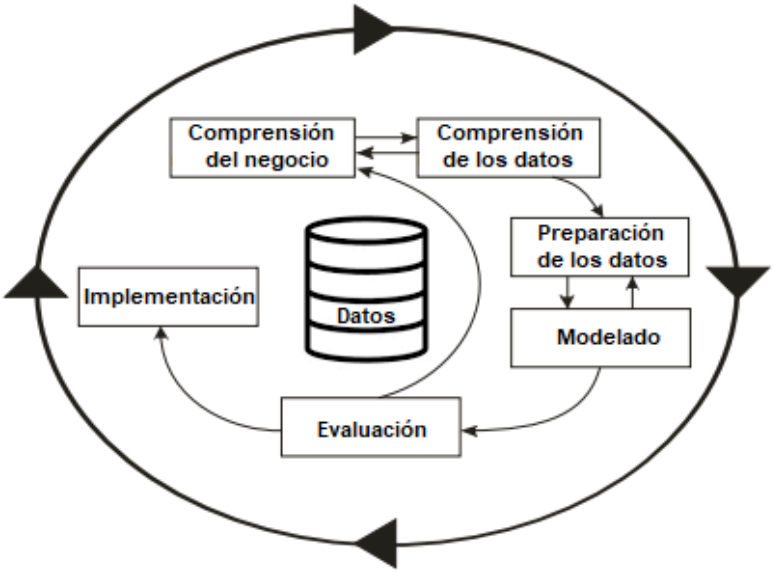
2.2.3. Metodología *CRISP-DM*

Una de las metodologías más destacada para la extracción del conocimiento a partir de los datos es *CRISP-DM*, *Cross-Industry Standard Proceess for Data Mining* (Chapman, y otros, 2000), la cual describe un modelo de proceso jerárquico en diferentes tareas que sugiere un ciclo de vida de un proyecto de minería de datos que consiste en seis fases dinámicas: comprensión del negocio, comprensión de los datos, preparación de datos, modelado, evaluación e implementación (figura 3).

Cada fase se estructura en tareas generales, éstas a su vez se proyectan en tareas específicas, en las cuales finalmente se describen las acciones que deben ser desarrolladas para situaciones definidas. Las fases de la metodología son:

- Fase 1. Comprensión del negocio o problema. Comprende los requisitos y objetivos del proyecto desde una perspectiva empresarial o institucional para convertirlos en objetivos técnicos y en un plan de proyecto, para lo cual es necesario comprender de manera completa el problema por resolver.

Figura 3. Ciclo de vida Metodología CRISP-DM.



Fuente: Modificado de Chapman, P. et al. (2000). CRISPDM 1.0 step-by-step data mining guide. Technical report, CRISP-DM.

- Fase 2. Comprensión de los datos. Corresponde a la recolección inicial de los datos para establecer un primer contacto con el problema; esta fase, junto con la fase 3 y la fase 4, demanda mayor esfuerzo y tiempo. Las principales tareas que se deben desarrollar en la fase de comprensión de los datos son: recolectar datos iniciales, describir los datos, explorar los datos y verificar la calidad de los datos.
- Fase 3. Preparación de los datos. Se usa para adaptarlos a la técnica de minería de datos, mediante la visualización de los datos y la búsqueda de relaciones entre las variables. Esta fase es la de modelado, ya que los datos requieren ser procesados de diferentes formas; por lo cual, las fases de preparación y modelado interactúan permanentemente
- Fase 4. Modelado. Corresponde a la selección de un modelo adecuado y específico; para ello se usan técnicas que cumplan los siguientes criterios
  - Ser apropiada para el problema.
  - Disponer de datos adecuados.
  - Cumplir con los requisitos del problema.

- Técnica adecuada para obtener un modelo.
- Conocimiento pleno de la técnica.

- Fase 5. Evaluación. Evalúa el modelo teniendo en cuenta el cumplimiento de los criterios de éxito del problema; para ello se emplean múltiples herramientas para la interpretación de los resultados. Si es válido, se procede a la explotación del modelo, que es el mantenimiento de la aplicación y la posible difusión de los resultados. Una vez que el modelo ha sido construido y validado, se transforma el conocimiento obtenido en acciones dentro del proceso de negocio; la retroalimentación generada por la monitorización y mantenimiento puede indicar si el modelo está siendo utilizado apropiadamente.
- Fase 6. Implementación. Es aquí donde el conocimiento obtenido se transforma en acciones dentro del proceso de negocio, ya sea observando el modelo y resultados, o aplicándolo a múltiples grupos de datos o como parte del proceso. Las tareas que se efectúan son: planear la implementación, monitorizar y mantener, informe final y revisar el proyecto

2.2.4. Mapas interactivos

Los mapas interactivos son una herramienta digital que permite presentar la información de manera dinámica, gráfica y selectiva; facilitando de esta manera, la comprensión visual y territorial sobre un tema determinado. Las posibilidades de acceso a la información y la transparencia ofrecida por los mapas interactivos permiten utilizarlos como medio para que la información sea completamente accesible y comprensible para toda persona (Medranda, Palacios, & Moromenacho, 2018).

2.3. MARCO TECNOLÓGICO

En el desarrollo del proyecto se utilizó software de libre distribución. A continuación, se presenta una breve descripción de las herramientas empleadas para su desarrollo.

2.3.1. Procesamiento y modelado de datos: *KNIME*[2]

KNIME (Konstanz Information Miner) es una plataforma analítica que permite el tratamiento de datos y desarrollo de modelos en un entorno visual. Está construido sobre la plataforma *Eclipse*, permitiendo manipular datos, analizarlos y crear flujos de trabajo. *KNIME* provee una serie de nodos, que encapsulan distintos tipos de algoritmos, y flechas que representan flujos de datos que se despliegan y combinan de manera gráfica e interactiva (KNIME A.G, 2020).

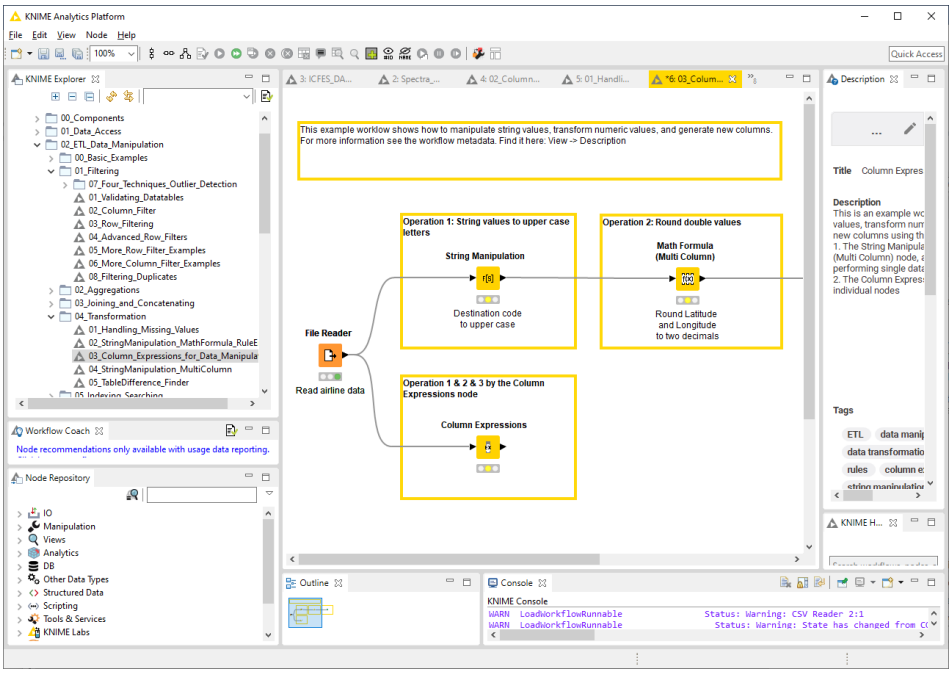
Los nodos implementan distintos tipos de acciones que pueden ejecutarse sobre una tabla de datos:

- Nodos de entrada y salida de datos: permite interactuar con bases de datos, archivos *csv*, libros de Excel, archivos de texto, entre otros.
- Manipulación de filas, columnas, etc., como muestreos, transformaciones, agrupaciones, etc.
- Visualización (histogramas, etc.).
- Creación de modelos estadísticos
- Minería de datos, como árboles de decisión, máquinas de vector soporte, regresiones, *clustering*, etc.

La interfaz gráfica de *KNIME* (ver figura 4) permite realizar el preprocesamiento de datos (ETL: extracción, transformación y carga), realizar el análisis y modelado de los datos y obtener una visualización.

[2] https://www.knime.com/

Figura 4. Interfaz herramienta KNIME



Fuente: Elaboración propia

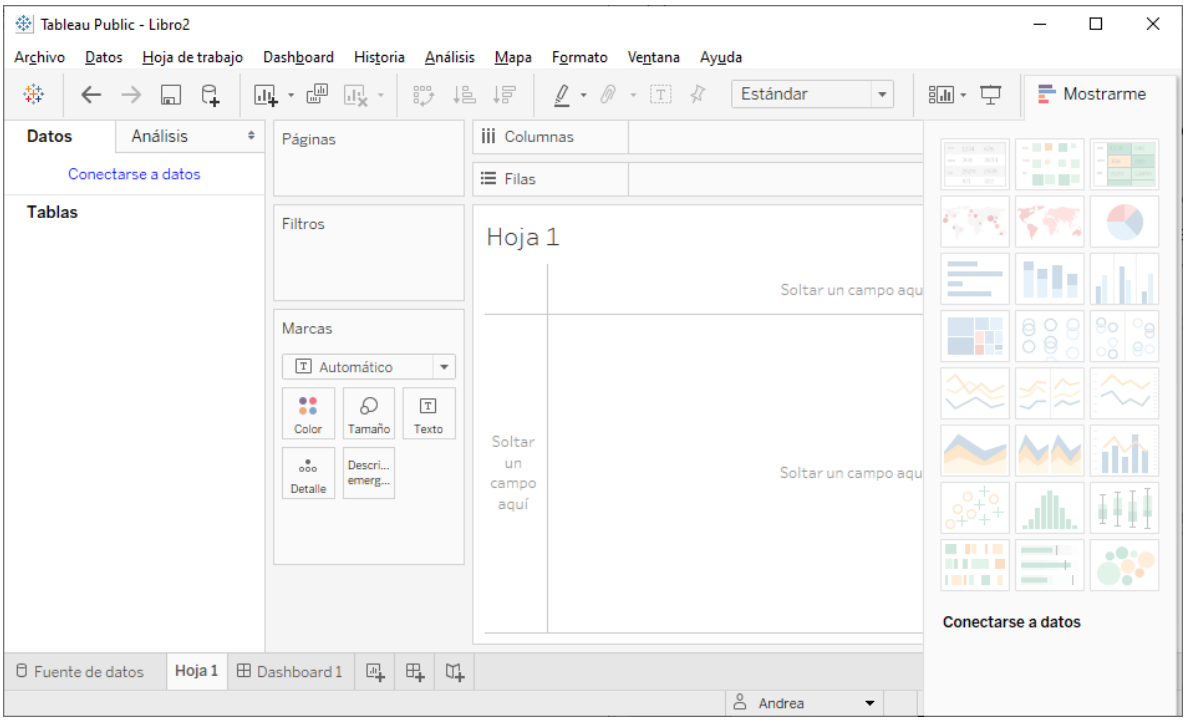
2.3.2. Herramienta de visualización: *Tableau*[3]

Aunque inicialmente se propuso *Google Data Studio* como herramienta de visualización, en el desarrollo del proyecto se evidenció las limitaciones de esta en cuanto a tamaños de archivos, tiempos de carga de datos y herramientas de formato gráfico; por esa razón se seleccionó *Tableau* en su versión pública.

Tableau Public es una herramienta de análisis y visualización de datos que permite trabajar con grandes volúmenes de datos a gran velocidad; a partir de diversos orígenes de datos (libros Excel, archivos *csv*, base de datos, entre otras) permitiendo su representación a través de múltiples gráficos y visualizaciones interactivas presentados de forma comprensible e intuitiva. Esta versión gratuita requiere la creación de una cuenta y que la información allí publicada pueda ser accedida por otros usuarios. Debido a que la información para el proyecto fue obtenida de datos abiertos, esto no supone inconveniente alguno.

[3] https://public.tableau.com/s/

Figura 5. Interfaz de la herramienta *Tableau Public*



Fuente: Elaboración propia

3. DISEÑO METODOLÓGICO

Para la realización del análisis de datos se aplicó una metodología *CRISP–DM* modificada. La metodología *CRISP-DM* describe un modelo de proceso jerárquico en un conjunto de tareas que proporcionan una descripción del ciclo de vida de un proyecto de minería de datos consistente diversas fases dinámicas. (Chapman, y otros, 2000)

Siguiendo esta metodología, se plantea que a partir de los resultados de pruebas académicas SABER 11 publicadas por el ICFES se desarrollen algunos análisis que permitan describir la situación académica colombiana y desarrollar un mapa interactivo de Colombia.

Las fases propuestas para el desarrollo del análisis son:

1. **Comprensión del negocio**

   Incluye la revisión de la información sobre las áreas de conocimiento evaluadas en las pruebas cuyos resultados son publicados en las bases de datos del ICFES: cálculos de puntaje, estructura y propósito de las pruebas.

2. **Entendimiento de los datos**

   En esta fase se desarrollan las actividades:

   - Recolección de datos.
   - Descripción del conjunto de datos: La descripción de datos incluye los atributos contenidos en los archivos de datos que se utilizarán.
   - Descripción de atributos: descripción de atributos numéricos y categóricos en las fuentes de datos.
   - Exploración de datos: permite analizar algunas variables e identificar características de algunas subpoblaciones y formular algunas preguntas de interés con relación al objetivo de minería.

- Verificación de la calidad de los datos: se identifican los valores de atributos que se consideran No Válidos, basados en las definiciones del diccionario de datos.
- Selección de datos: Se seleccionan los atributos para preprocesamiento y modelamiento
- Limpieza de datos: Se alinean a los requerimientos del algoritmo a aplicar (datos faltantes, valores que no han sido identificados previamente).

3. Selección de técnica de modelado

Se selecciona una técnica de modelado descriptiva de datos que pueda aplicarse a datos suministrados.

4. Modelamiento de datos

Donde se identifican las variables o parámetros característicos de los grupos de datos analizados.

5. Desarrollo de herramienta para visualización de la información

Desarrollo de la herramienta que contenga el mapa interactivo con la información de los resultados de las pruebas saber para el periodo 2015-2019.

6. Análisis de la situación académica colombiana

En base a la información estudiada, se realiza el análisis de la situación académica colombiana, por regiones y/o departamentos resaltando oportunidades de mejoramiento, tendencias positivas e información intrínseca que se descubra entre los datos estudiados.

4. PROPUESTA Y ANÁLISIS DE RESULTADOS

A continuación, se detallan las fases de la metodología implementada y sus principales hallazgos.

4.1. COMPRENSIÓN DEL NEGOCIO

El ICFES es la entidad encargada de gestionar la calidad educativa en el país; para cumplir con su objetivo, desarrolla diversas evaluaciones a lo largo de la vida académica de los estudiantes, desde su formación básica primaria hasta la formación profesional.

Una de esas evaluaciones es el examen Saber 11, el cual es una prueba estandarizada que se aplica a los estudiantes de último año de educación media. Inicialmente fue creado con el propósito de apoyar los procesos de admisión de las instituciones de educación superior, siendo un requisito para el ingreso a cualquier programa de pregrado. Sin embargo, debido a su orientación a la evaluación por competencias, es un examen que permite cuantificar los aportes de las instituciones educativas en términos de aprendizaje, permite recopilar información para la construcción de indicadores de calidad de la educación y ofrece información como referente estratégico para el establecimiento de políticas educativas, nacionales, territoriales e institucionales, entre otras (ver figura 6). Por estas razones, fue seleccionada la prueba Saber 11 para el análisis de la situación académica colombiana.

Debido a la existencia de dos calendarios académicos en el país, el examen Saber 11 se aplica semestralmente. En el primer semestre presentan el examen los estudiantes de colegios de calendario B y en el segundo semestre los estudiantes del calendario A; por esta razón, en el repositorio de información se encuentran dos archivos de reporte de resultados por año.

Con el objetivo de consolidar el Sistema Nacional de Evaluación Estandarizada (SNEE) que permitiera la alineación de todos los exámenes que lo conforman, a partir del segundo semestre del 2014 se reestructuró el examen, para evaluar cinco

pruebas genéricas en total: matemáticas, lectura crítica, ciencias naturales, sociales-y-ciudadanas e inglés.

Figura 6. Objetivos del examen Saber 11.

Fuente: ICFES. (2019). Informe nal. de resultados del examen saber 11° 2018.

Los resultados contienen puntajes del examinando en cada una de las cinco pruebas genéricas en una escala fijada con promedio 50 y desviación estándar 10 y un puntaje global, construido a partir de un promedio ponderado de los puntajes en las cinco pruebas genéricas bajo la ecuación 1.

$PG=5*IG$ ec. 1

Donde:

$$IG = \frac{3*\text{MATEMÁTICAS}+3*\text{LECTURA}+3*\text{CIENCIAS}+3*\text{SOCIALES}+1*\text{INGLÉS}}{13}$$

La tabla 2 contiene las principales características de las escalas de los puntajes resultantes para cada prueba y el puntaje global.

Tabla 2. Escalas de puntajes en el examen Saber 11

Puntaje	Media	Desviación Estándar	Mínimo	Máximo
Por prueba	50	10	0	100
Global	250	50	0	500

Fuente: Elaboración propia

De forma complementaria a la composición del examen, la prueba SABER 11 incluye un cuestionario que indaga por información personal, académica y socioeconómica de los estudiantes y en algunas versiones del examen se aplican cuestionarios como los de antecedentes escolares y expectativas sobre educación superior a una muestra aleatoria de la población. (ICFES, 2018b).

- **Información personal:** Este módulo indaga por aspectos como el género del estudiante, pertenencia a una etnia, lugar de residencia, entre otros.

- Información Académica y de citación: Este módulo indaga por aspectos de la institución educativa a la que pertenece el estudiante tales como calendario, jornada, naturaleza oficial o privada, área de ubicación, entre otros.

- Información socioeconómica: Este módulo indaga por aspectos familiares como el nivel educativo de los padres, su ocupación, servicios con los que cuenta el hogar, estrato de la vivienda, número de personas en la familia, tenencia de computador, acceso a servicio de internet, entre otros.

Como parte de su estrategia, el ICFES se está promoviendo el uso de los resultados de los exámenes aplicados en los diversos niveles educativos para mejorar la comprensión de avance académico y plantear estrategias de mejoramiento de la calidad, cobertura y apoyo a la educación.

Con el desarrollo de este proyecto, a los resultados de los exámenes de la prueba Saber 11 aplicados a partir del primer semestre del 2015 hasta el segundo semestre del 2019 se le realizará un análisis descriptivo y un procesamiento con el fin de lograr un mejor entendimiento de los resultados en un contexto regional y nacional.

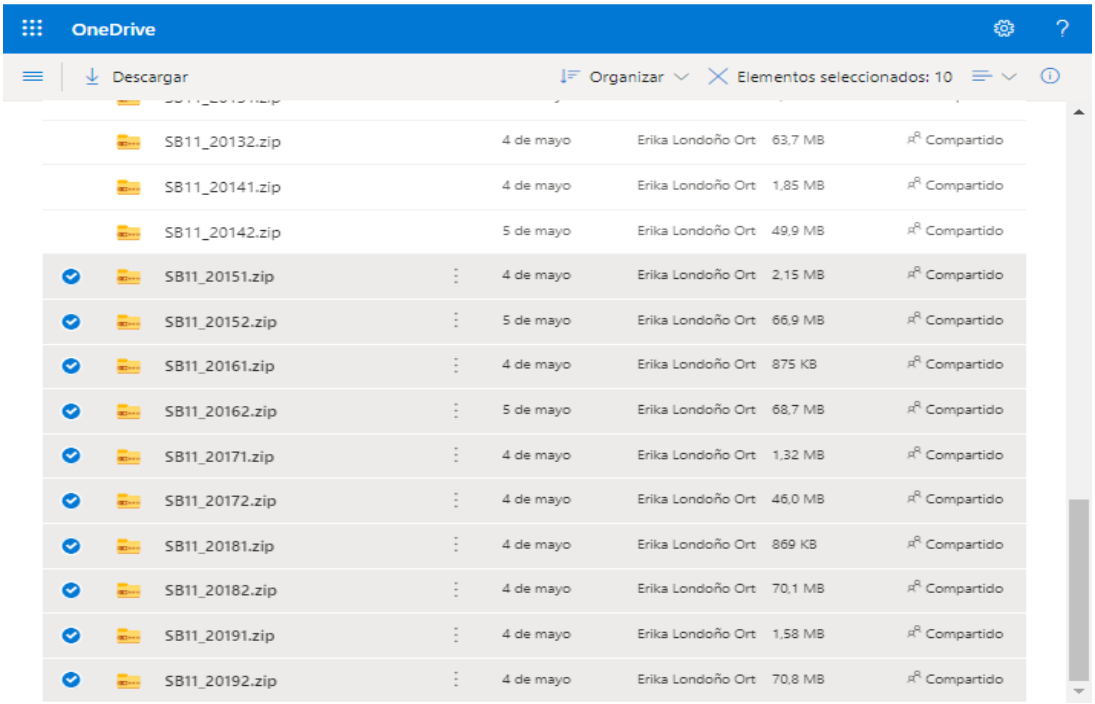
4.2. ENTENDIMIENTO DE LOS DATOS

En esta fase se realizó la captura de los datos a analizar para identificar su estructura, realizar su descripción, identificar problemas de calidad en los mismos y detectar subconjuntos de datos que pudieran ser útiles para formular hipótesis específicas que se validan posteriormente con el análisis. A continuación, se describen las diferentes actividades llevadas a cabo en esta fase.

4.2.1. Recolección de datos

Los datos para el desarrollo del proyecto fueron obtenidos del repositorio de información del ICFES. Se consultaron los resultados individuales de las pruebas Saber 11 a partir del primer semestre del 2015 hasta el segundo semestre de 2019. Por cada año se reportan dos archivos, correspondientes a cada semestre del año. En total se descargaron 10 archivos de texto que se presentan en la figura 7.

Figura 7. Archivos descargados del repositorio de información

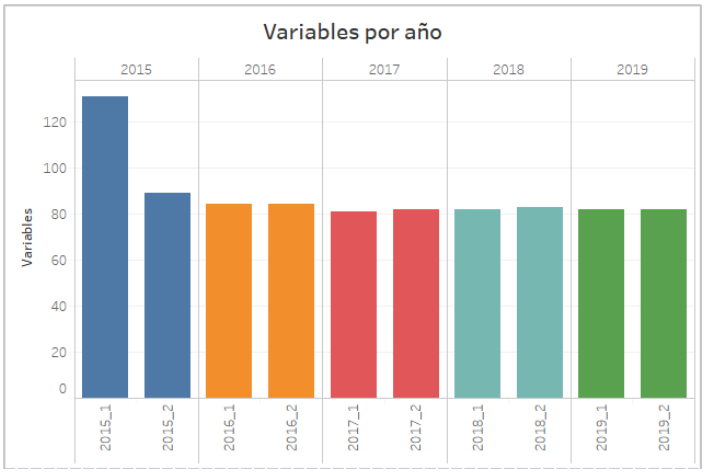


Fuente: Elaboración propia

4.2.2. Descripción del conjunto de datos

La descripción de datos corresponde a los atributos contenidos en los archivos a utilizar para el análisis. En los periodos a evaluar, el número de atributos oscila entre 81 y 131, como se presenta en la figura 8.

Figura 8. Atributos evaluados por año-periodo



Fuente: Elaboración propia

Hallazgos en la revisión preliminar de atributos:

1. La dimensionalidad de los atributos es multivariada con datos categóricos y numéricos.
2. Algunos atributos, pueden tomar valores diferentes dependiendo del año de estudio, como los que se muestran a continuación:
    - Atributo FAMI_PERSONASHOGAR: Para los años 2015-2016 puede tomar los valores independientes de: *Una, Dos, Tres, Cuatro, Cinco, Seis, Siete, Ocho, Nueve, Diez, Once, Doce o más.* Para los años 2017 a 2019 toma rangos de valores: *1 a 2, 3 a 4, 5 a 6, 7 a 8, 9 o más.*
    - Atributo FAMI_CUARTOSHOGAR: Para los años 2015-2016 puede tomar los valores de *Un**a,** Dos, Tres, Cuatro, Cinco, Seis, Siete, Ocho, Nueve, Diez o más*. Para los años 2017 a 2019 restringen un poco más los valores: *Un**o**, Dos, Tres, Cuatro, Cinco, Seis o más.*

Para que estos atributos puedan ser comparables en el tiempo, es necesario unificar sus valores; y debido a que los periodos 2017-2019 incluyen los valores más agrupados, es conveniente transformar los valores de 2015-2016 a este formato.

3. Existen atributos que, aunque tengan nombres diferentes para algunos periodos, indican la misma característica. Por ejemplo, los atributos FAMI_TIENETELEVISOR (presente en los resultados del primer periodo del 2015 y el primer periodo del 2016) y FAMI_TIENESERVICIOTV (presente en los demás periodos) indican si en el hogar del evaluado hay servicio de tv y pueden tomar los valores de SI y NO. Para que sean comparables, se debe unificar el nombre del atributo.

4. Existen atributos que se encuentran solo para algunos años, y para otros años existen atributos similares que, aunque tengan nombre diferente y puedan tomar otros valores, permiten identificar una característica del evaluado, por ejemplo:

   - El atributo ESTU_TRABAJAACTUALMENTE (presente en los años 2015-2016) indica si el evaluado trabaja y cuántas horas a la semana; puede tomar los valores de: "*No", "Si menos de 20 horas", "Si 20 horas o más";* el atributo ESTU_HORASSEMANATRABAJA (presente en los años 2017-2019) indica la cantidad de horas que el estudiante trabaja a la semana y puede tomar los valores de: "*0", "menos de 10", "entre 11 y 20", "entre 21 y 30", "más de 30".* Estos dos atributos se pueden unificar convirtiendo los valores de ESTU_HORASSEMANATRABAJA a los valores del parámetro ESTU_TRABAJAACTUALMENTE. Es decir, si el estudiante trabaja *"0"* horas a la semana, toma el valor de *"No"*; si trabaja *"menos de 10 horas"* se asigna el valor "*Si menos de 20 horas",* y así sucesivamente.

   - El parámetro ESTU_RECIBESALARIO (presente en los años 2015-2016) indica si el estudiante recibe o no remuneración y puede tomar los valores *"Si"* y *"No",* el parámetro ESTU_TIPOREMUNERACION (presente en los años 2017-2019) indica si recibe o no remuneración y de qué tipo, tomando los valores de *"No", "Si en efectivo", "Si en especie", "Si en efectivo y*

especie", para que los años sean comparables se pueden convertir los valores de ESTU_TIPOREMUNERACION en *"Si", "No".*

- Los atributos FAMI_OCUPACIONPADRE (presente en los años 2015-2016) y FAMI_LABORPADRE (presente en los años 2017-2019) indican el trabajo o labor que desempeña el padre del evaluado. El primero puede tomar 12 valores diferentes y el segundo 13. Para que estos valores sean comparables en el tiempo, se agruparon en nuevas categorías. Los atributos FAMI_OCUPACIONMADRE (presente en los años 2015-2016) y FAMI_LABORMADRE presentan esta situación y se manejan de igual forma. La tabla A1 del anexo A presenta esta clasificación.

5. Después de revisar los parámetros comunes y los semejantes (que se pueden unificar) se obtuvo un total de 60 atributos. Los campos que no cumplen esta condición fueron omitidos. La tabla A2 del anexo A presenta este listado.

Los atributos preliminares se encuentran clasificados en seis grupos: información personal del evaluado (9), información de contacto (4), información socioeconómica (17), información del colegio (17), datos de citación del examen (5) y resultados (8).

Las tablas 3-8 describen los atributos correspondientes y la tabla A3 del anexo A detalla los valores que pueden tomar estos campos para los periodos evaluados.

Tabla 3. Atributos: Información personal del evaluado

Campo	Descripción
ESTU_TIPODOCUMENTO	Tipo de Documento
ESTU_NACIONALIDAD	Nacionalidad
ESTU_GENERO	Género del evaluado
ESTU_FECHANACIMIENTO	Fecha de Nacimiento
PERIODO	Periodo de presentación del examen
ESTU_CONSECUTIVO	Id público del inscrito (SB11)
ESTU_ESTUDIANTE	Indica si el inscrito realizó la inscripción por medio de un colegio o de forma particular
ESTU_PAIS_RESIDE	País donde reside actualmente
ESTU_ETNIA	Grupo étnico minoritario al que pertenece

Fuente: Elaboración propia

Tabla 4. Atributos: Información de contacto

Campo	Descripción
ESTU_DEPTO_RESIDE	Departamento de residencia
ESTU_COD_RESIDE_DEPTO	Código Dane del departamento de residencia
ESTU_MCPIO_RESIDE	Municipio de Residencia
ESTU_COD_RESIDE_MCPIO	Código Dane del municipio de residencia

Fuente: Elaboración propia

Tabla 5. Atributos: Información socioeconómica

Campo	Descripción
FAMI_ESTRATOVIVIENDA	Estrato socioeconómico de la vivienda según recibo de energía eléctrica.
FAMI_PERSONASHOGAR	Personas que conforman el hogar, incluyendo el estudiante
FAMI_CUARTOSHOGAR	Número de cuartos donde duermen las personas del hogar
FAMI_EDUCACIONPADRE	Nivel educativo más alto alcanzado por el padre
FAMI_EDUCACIONMADRE	Nivel educativo más alto alcanzado por la madre
FAMI_TRABAJOLABORPADRE	Labor que realizó el padre durante la mayor parte del último año
FAMI_TRABAJOLABORMADRE	Labor que realizó la madre durante la mayor parte del último año
FAMI_TIENEINTERNET	¿El hogar cuenta con servicio o conexión a internet?
FAMI_TIENESERVICIOTV	¿El hogar cuenta con servicio cerrado de televisión?
FAMI_TIENECOMPUTADOR	¿El hogar posee Computador?
FAMI_TIENELAVADORA	¿El hogar posee Máquina lavadora de ropa?
FAMI_TIENEAUTOMOVIL	¿El hogar posee Automóvil particular?
FAMI_NUMLIBROS	Cantidad de libros físicos o electrónicos hay en el hogar excluyendo periódicos, revistas, directorios telefónicos y libros del colegio.
ESTU_INSE_INDIVIDUAL	Índice Socioeconómico del evaluado
ESTU_NSE_INDIVIDUAL	Nivel Socioeconómico del evaluado
ESTU_TRABAJAACTUALMENTE	¿Trabaja actualmente?
ESTU_RECIBESALARIO	¿Recibe remuneración por trabajar?

Fuente: Elaboración propia

Tabla 6. Atributos: Información del colegio

Campo	Descripción
COLE_CODIGO_ICFES	Código ICFES de la sede-jornada
COLE_COD_DANE_ESTABLECIMIENTO	Código Dane del colegio
COLE_NOMBRE_ESTABLECIMIENTO	Nombre del colegio
COLE_GENERO	Género de la población del colegio
COLE_NATURALEZA	Naturaleza del colegio
COLE_CALENDARIO	Calendario académico del colegio
COLE_BILINGUE	Indica si el colegio es bilingüe o no
COLE_CARACTER	Carácter del colegio
COLE_COD_DANE_SEDE	Código Dane de la Sede
COLE_NOMBRE_SEDE	Nombre de la Sede
COLE_SEDE_PRINCIPAL	Indica si es la sede principal del colegio
COLE_AREA_UBICACION	Área de ubicación de la Sede
COLE_JORNADA	Jornada de la Sede
COLE_COD_MCPIO_UBICACION	Código Dane del municipio donde está ubicada la Sede
COLE_MCPIO_UBICACION	Nombre del municipio donde está ubicada la Sede
COLE_COD_DEPTO_UBICACION	Código Dane del departamento de la Sede
COLE_DEPTO_UBICACION	Nombre del departamento donde está ubicada la Sede

Fuente: Elaboración propia

Tabla 7. Atributos: Datos de citación del examen

Campo	Descripción
ESTU_PRIVADO_LIBERTAD	Indica si el estudiante se encuentra privado de la libertad
ESTU_COD_MCPIO_PRESENTACION	Código Dane del municipio presentación del examen
ESTU_MCPIO_PRESENTACION	Municipio de presentación del examen
ESTU_DEPTO_PRESENTACION	Departamento de presentación del examen
ESTU_COD_DEPTO_PRESENTACION	Código Dane del departamento de presentación del examen

Fuente: Elaboración propia

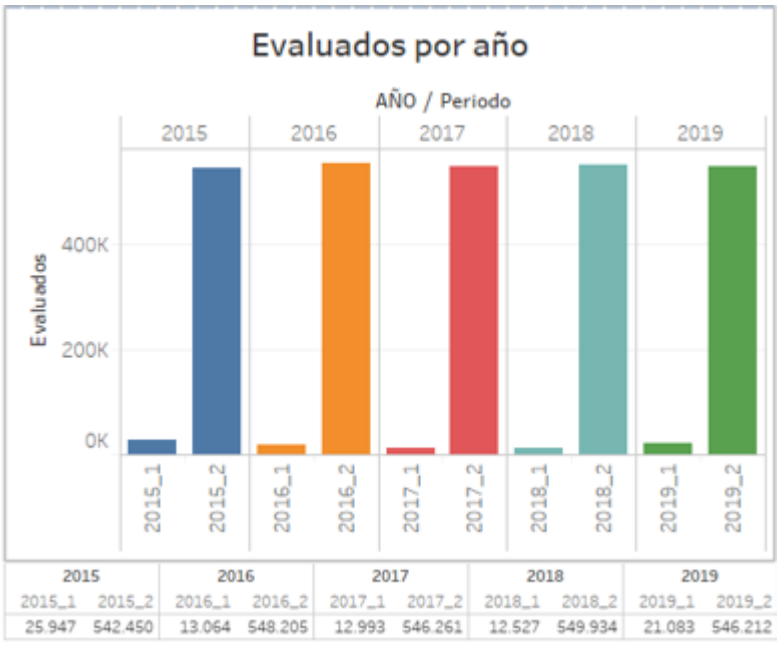
Tabla 8. Atributos: Resultados

Campo	Descripción
PUNT_LECTURA_CRITICA	Puntaje en lectura crítica
PUNT_MATEMATICAS	Puntaje en matemáticas
PUNT_C_NATURALES	Puntaje en ciencias naturales
PUNT_SOCIALES_CIUDADANAS	Puntaje sociales y ciudadanas
PUNT_INGLES	Puntaje inglés
DESEMP_INGLES	Desempeño en inglés
PUNT_GLOBAL	Puntaje total obtenido
ESTU_ESTADOINVESTIGACION	Estado de los resultados de los evaluados

Fuente: Elaboración propia

El número total de registros de resultados es de 2’818.676; los cuales se encuentran distribuidos en 5 años y 10 periodos como se presenta en la figura 9. En esta gráfica se puede apreciar que la cantidad de estudiantes que presenta el examen en el primer semestre es considerablemente inferior (menos del 5%) a la cantidad estudiantes que presentan la prueba en el segundo período, lo que confirma que la mayoría de los establecimientos educativos en Colombia maneja calendario A.

Figura 9. Evaluados por año-periodo



Fuente: Elaboración propia

4.2.3. Descripción de atributos

Se identificaron algunos atributos cuya descripción detallada se considera innecesaria, ya sea porque aportan información redundante, no poseen información o son valores consecutivos y fueron omitidos. La tabla 9 presenta estos atributos omitidos.

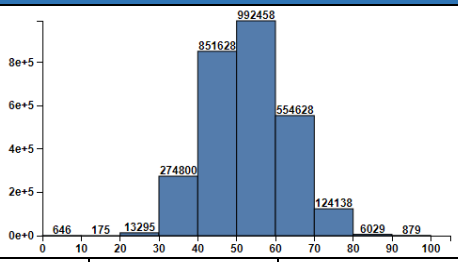
Tabla 9. Atributos omitidos

Atributo	Razón de omisión
ESTU_CONSECUTVO	Número consecutivo identificador del estudiante, debido a su naturaleza anónima la información que posee es nula.
ESTU_ESTUDIANTE	Es un parámetro que en todas las instancias tiene el mismo valor "ESTUDIANTE" y no aporta información alguna.
ESTU_COD_RESIDE_DPTO EST_COD_RESIDE_MPIO EST_COD_DANE_ESTABLECIMIENTO EST_COD_DANE_SEDE EST_COD_MPIO_UBICACION COLE_COD_DEPTO_UBICACION ESTU_COD_MPIO_PRESENTACION ESTU_COD_DEPTO_PRESENTACION	Existen dos formas de referirse al departamento, municipio, establecimiento y sede del establecimiento: con códigos y con nombres. Los códigos son complemento de los nombres y para el análisis descriptivo no aportan información.

Fuente: Elaboración propia

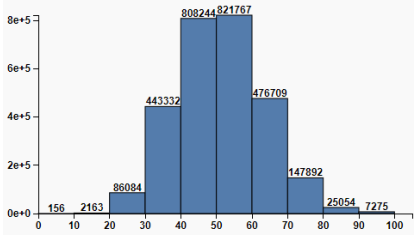
Las dimensiones del conjunto de datos analizado son multivariadas, con datos categóricos nominales y numéricos. Para describir el conjunto de datos numéricos se utilizaron medidas de tendencia central y de dispersión, para cada año de estudio. Adicionalmente se realizaron los histogramas de frecuencias de la totalidad de los datos. Las tablas 10–16 presentan los resultados obtenidos.

Tabla 10. Atributo numérico. Puntaje lectura crítica

PUNT_LECTURA_CRITICA					
					
MEDIDA / AÑO	**2015**	**2016**	**2017**	**2018**	**2019**
RANGO	0 - 100	0 - 100	0 - 100	0 - 100	0 - 100
MEDIA	49.921	52.772	53.399	52.822	52.372
σ	9.535	9.704	9.889	10.251	10.664
VARIANZA	90.907	94.162	97.8	105.077	113.717
FALTANTES	0	0	0	0	0
En el área de lectura crítica se aprecia inicialmente un aumento constante del promedio a nivel nacional entre los años 2015 y 2017; sin embargo, a partir del 2018 se identifica una leve disminución en los puntajes obtenidos.					

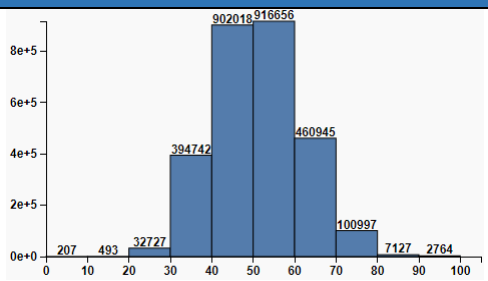
Fuente: Elaboración propia

Tabla 11. Atributo numérico. Puntaje matemáticas

PUNT_MATEMÁTICAS					
					
MEDIDA /AÑO	**2015**	**2016**	**2017**	**2018**	**2019**
RANGO	0 - 100	0 - 100	0 - 100	0 - 100	0 - 100
MEDIA	50.358	51.18	50.348	50.494	50.84
σ	12.277	12.07	12.156	12.112	12.176
VARIANZA	150.718	145.695	147.76	146.69	148.254
FALTANTES	0	0	0	0	0
En el área de matemáticas se ve un leve aumento del promedio en el 2016, sin embargo, al año siguiente se observa una disminución de los puntajes que contrasta con los obtenidos en el 2015; a partir del 2016 se identifica un aumento constante (leve) en los puntajes obtenidos, aunque no alcanzan los niveles logrados en 2016.					

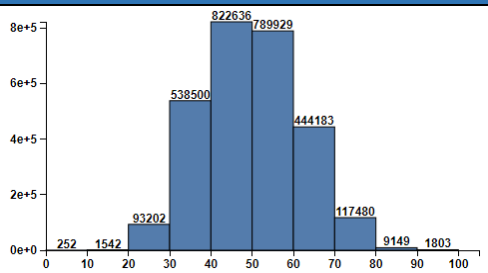
Fuente: Elaboración propia

Tabla 12. Atributo numérico. Puntaje ciencias naturales

PUNT_CIENCIAS NATURALES					
					
MEDIDA /AÑO	**2015**	**2016**	**2017**	**2018**	**2019**
RANGO	0 - 100	0 - 100	0 - 100	0 - 100	0 - 100
MEDIA	50.276	52.883	51.487	49.822	48.48
σ	10.405	9.895	10.201	10.64	10.952
VARIANZA	108.26	97.906	104.065	113.203	119.953
FALTANTES	0	0	0	0	0
Al igual que matemáticas, en el área de ciencias naturales se identificó un leve aumento del promedio en el 2016, sin embargo, a partir del año siguiente se observa una disminución de los puntajes que continúa hasta el último año evaluado.					

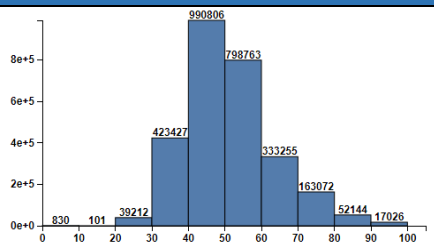
Fuente: Elaboración propia

Tabla 13. Atributo numérico. Puntaje Sociales y Ciudadanas

PUNT_SOCIALES Y CIUDADANAS					
					
MEDIDA /AÑO	**2015**	**2016**	**2017**	**2018**	**2019**
RANGO	0 - 100	0 - 100	0 - 100	0 - 100	0 - 100
MEDIA	50.049	50.841	50.498	48.503	46.522
σ	11.425	10.969	11.264	11.947	12.331
VARIANZA	130.537	120.309	126.87	142.723	152.048
FALTANTES	0	0	0	0	0
Al igual que en ciencias naturales, en el área de sociales-y-ciudadanas se identificó un leve aumento del promedio en el 2016, sin embargo, a partir del año siguiente se observa una disminución de los puntajes que continúa hasta el último año evaluado.					

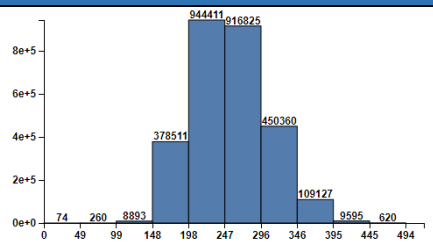
Fuente: Elaboración propia

Tabla 14. Atributo numérico. Puntaje Inglés

PUNT_INGLÉS					
					
MEDIDA /AÑO	**2015**	**2016**	**2017**	**2018**	**2019**
RANGO	0 - 100	0 - 100	0 - 100	0 - 100	0 - 100
MEDIA	50.728	52.431	49.993	51.045	48.863
σ	11.468	12.195	11.994	11.957	13.021
VARIANZA	131.516	148.718	143.852	142.962	169.543
FALTANTES	0	0	0	0	19
El comportamiento de los resultados del área de ingles indica un aumento del puntaje para el año 2016; sin embargo, a diferencia de las otras áreas, en los años siguientes no se observa una tendencia, ya que va aumentando y disminuyendo.					

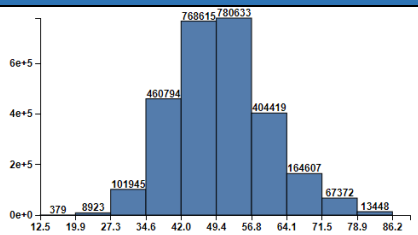
Fuente: Elaboración propia

Tabla 15. Atributo numérico. Puntaje Global

PUNT_GLOBAL					
					
MEDIDA /AÑO	**2015**	**2016**	**2017**	**2018**	**2019**
RANGO	0 - 492	0 - 494	0 - 476	0 - 478	0 - 477
MEDIA	250.977	259.792	256.611	252.296	247.501
σ	48.496	48.647	49.34	51.071	52.441
VARIANZA	2351.876	2366.562	2434.451	2608.293	2750.055
FALTANTES	0	0	0	0	0
Debido a que el puntaje global es calculado en base a los puntajes de las áreas individuales, en el puntaje global se observa el mismo comportamiento de las áreas de Ciencias Naturales y Sociales-y-Ciudadanas en donde se observó un aumento en los puntajes del 2016, pero en los años siguientes los promedios van disminuyendo.					

Fuente: Elaboración propia

Tabla 16. Atributo numérico. INSE del evaluado

ESTU_INSE_INDIVIDUAL					
					
MEDIDA /AÑO	**2015**	**2016**	**2017**	**2018**	**2019**
RANGO	16.9-86.2	17- 86.2	14.4-84.3	12.5-84.8	14.5 -84.8
MEDIA	50.053	49.834	50.639	50.31	50.614
σ	10.075	9.824	9.309	10.004	10.022
VARIANZA	101.502	96.51	86.651	100.087	100.45
FALTANTES	1159	2684	7321	18486	17891
El INSE (índice de nivel socioeconómico) es un índice calculado en base a las características socioeconómicas del evaluado. La poca variación de este índice en los periodos de estudio indica que el perfil socioeconómico de los evaluados en la prueba SABER 11 se mantiene constante.					

Fuente: Elaboración propia

Tabla 17. Descripción de atributos categóricos 2015-2019

PARAMETRO	VAL.	MODA	MISSINGS
ESTU_TIPODOCUMENTO	10	TI (80%)	0
ESTU_NACIONALIDAD	85	COLOMBIA (99.9%)	0
ESTU_GENERO	2	F (54.3%)	2668 (0.1%)
ESTU_FECHANACIMIENTO (AÑO DE NACIMIENTO) *	131	2000 (18.8%)	3
PERIODO	10	2018_2 (19.5%)	0
ESTU_PAIS_RESIDE	85	COLOMBIA (99.9%)	0
ESTU_ETNIA	20	COMUNIDAD AFRODESCENDIENT (3.4%)	2619081 (92.9%)
ESTU_DEPTO_RESIDE	35	Bogotá (16.5%)	877 (0.03%)
ESTU_MCPIO_RESIDE	>1000	Bogotá	877 (0.03%)

PARAMETRO	VAL.	MODA	MISSINGS
FAMI_ESTRATOVIVIENDA	7	ESTRATO 1 (35.5%)	105862 (3.8%)
FAMI_PERSONASHOGAR	5	3 a 4 (47.5%)	51183 (1.8%)
FAMI_CUARTOSHOGAR	6	TRES (39.6%)	50561 (1.8%)
FAMI_EDUCACIONPADRE	12	Secundaria (Bachillerato) completa (22.3%)	96355 (3.4%)
FAMI_EDUCACIONMADRE	12	Secundaria (Bachillerato) completa (25.7%)	95983 (3.4%)
FAMI_TRABAJOLABORPADRE	13	Empleado (vendedor-operario-obrero-conductor-limpieza-mtto-seguridad) (23.2%)	56589 (2%)
FAMI_TRABAJOLABORMADRE	13	Trabaja en el hogar o no trabaja o no estudia (44.2%)	52811 (1.9%)
FAMI_TIENEINTERNET	2	Si (55.1%)	97958 (3.5%)
FAMI_TIENESERVICIOTV	2	Si (70%)	102422 (3.6%)
FAMI_TIENECOMPUTADOR	2	Si (58.2%)	50432 (1.8%)
FAMI_TIENELAVADORA	2	Si (71.3%)	49552 (1.8%)
FAMI_TIENEAUTOMOVIL	2	No (76.84%)	54283 (1.9%)
FAMI_NUMLIBROS	4	0 a 10 (43.19%)	125856 (4.46%)
ESTU_NSE_INDIVIDUAL	4	NSE2 (37.3%)	47541 (1.7%)
ESTU_TRABAJAACTUALMEN_TE	4	No (76.87%)	50005 (1.8%)
ESTU_RECIBESALARIO	2	No (41.8%)	1078292 (38.3%)
COLE_NOMBRE_ESTABLECIMIENTO	>10000	---	0
COLE_GENERO	3	Mixto (96.3%)	0
COLE_NATURALEZA	2	Oficial	0

PARAMETRO	VAL.	MODA	MISSINGS
		(73.2%)	
COLE_CALENDARIO	3	A (96.9%)	0
COLE_BILINGUE	2	No (84.3%)	398008 (14%)
COLE_CARACTER	4	Académico (53.9%)	45995 (1.6%)
COLE_NOMBRE_SEDE	>10000	---	0
COLE_SEDE_PRINCIPAL	2	S (97%)	0
COLE_AREA_UBICACION	2	Urbano (85.2%)	0
COLE_JORNADA	6	Mañana (49.7%)	0
COLE_MCPIO_UBICACION	>1000	Bogotá	0
COLE_DEPTO_UBICACION	34	Bogotá (16.3%)	0
ESTU_PRIVADO_LIBERTAD	2	N (99.98%)	0
ESTU_MCPIO_PRESENTACION	483	Bogotá (16.5%)	86 (0.003%)
ESTU_DEPTO_PRESENTACION	34	Bogotá (16.5%)	86 (0.003%)
DESEMP_INGLES	5	A- (44.9%)	0
ESTU_ESTADOINVESTIGACION	4	PUBLICAR (99.88%)	0

Fuente: Elaboración propia

De los valores obtenidos en la tabla anterior, inicialmente se podría pensar que la mayoría de evaluados en las pruebas SABER 11 son estudiantes menores de edad (documento de identidad más común es tarjeta de identidad). Adicionalmente se observa que más del 99% de evaluados son colombianos y residen en el país.

En Bogotá están ubicados la mayor parte de evaluados de las pruebas. La mayoría de los estudiantes pertenecen al estrato 1 y al nivel socioeconómico 2. Más del 50% de los estudiantes tiene computador y servicio de internet. La mayoría de los colegios son de carácter mixto con calendario A y de carácter académico.

El atributo ESTU_ETNIA contiene el 92% de los datos totales como faltantes (*missings*), sin embargo, estos registros se consideran no nulos debido a que, para algunos periodos en estudio, el vacío indica que no pertenece a ninguna etnia.

De forma similar, COLE_BILINGUE tiene 14% de faltantes, esto es debido a que en algunos periodos cuando el colegio no es bilingüe no se llenaba ningún dato. Por esta razón los datos faltantes indican que el colegio no es bilingüe.

Adicionalmente, ESTU_RECIBESALARIO tiene 38% de faltantes debido a que en algunos periodos cuando ESTU_TRABAJAACTUALMENTE="no" no se llenaba el parámetro ESTU_RECIBESALARIO.

El atributo FECHA_NACIMIENTO se evaluó únicamente para el año de nacimiento. Es importante mencionar que al revisar los años de nacimiento se encuentran fechas que van desde el año 1 hasta el año 2099 y que más adelante tendrán que ser revisados.

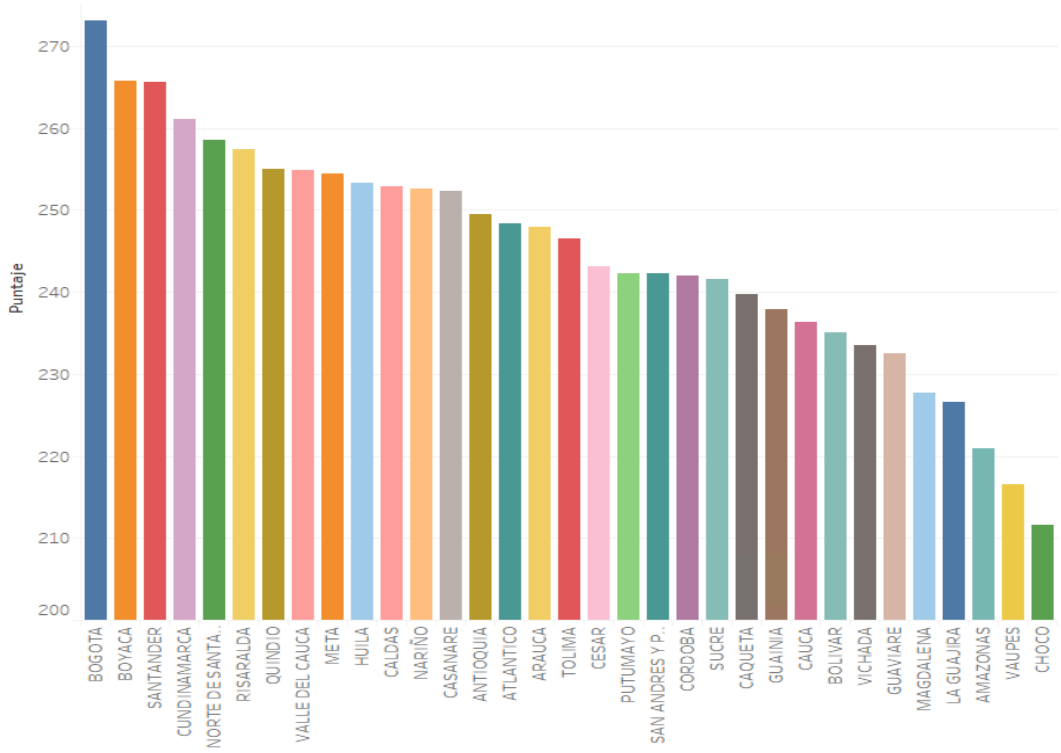
4.2.4. Exploración de datos

En esta etapa se realizó una exploración inicial del conjunto de datos; se plantearon algunos interrogantes y se analizaron algunas variables para explorar la situación académica en el país.

1. ¿Qué regiones o departamentos obtienen los mejores resultados y los resultados menos favorables?

Para resolver esta inquietud se estudiaron los puntajes promedio obtenidos por los departamentos en el periodo de tiempo analizado (2015-2019). En la figura 10 se observa que los mejores resultados se encuentran en la zona andina en Bogotá, Boyacá y Santander. En contraste, los departamentos de Amazonas, Vaupés y Chocó son los que presentan resultados menos favorables.

Figura 10. Puntaje global promedio por departamento periodo 2015-2019

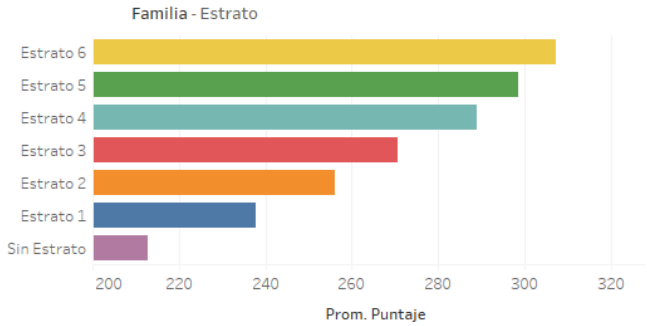


Fuente: Elaboración propia

3. ¿El estrato de los estudiantes influye en su desempeño académico?

Para resolver esta inquietud se estudiaron los puntajes globales promedio obtenidos por cada estrato en el periodo 2015-2019. En la figura 11 se evidencia que el estrato influye positivamente en el logro de mejores resultados. A mayor estrato mejores son los resultados obtenidos.

Figura 11. Desempeño de estudiantes por estrato

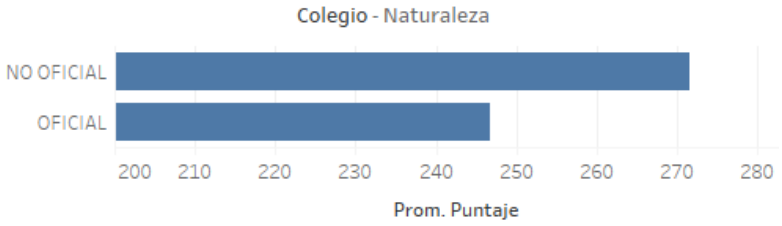


Fuente: Elaboración propia

4. ¿Las características de la institución educativa influyen en el desempeño de los estudiantes?

Se contemplaron dos características de la institución educativa: naturaleza y área de ubicación. La figura 12 presenta el puntaje promedio obtenido por los colegios oficiales y no oficiales, donde se observa que los colegios de naturaleza no oficial alcanzan puntajes globales promedio superiores a las instituciones oficiales.

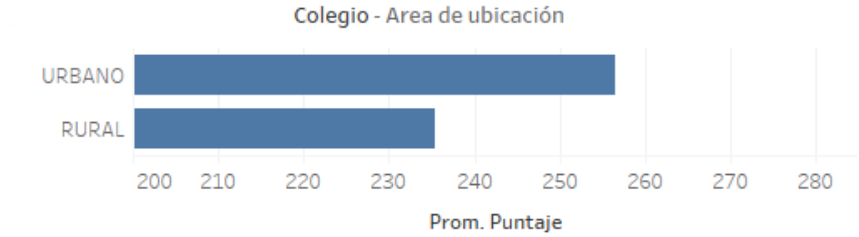
Figura 12. Desempeño de estudiantes por naturaleza del colegio



Fuente: Elaboración propia

La figura 13 presenta el puntaje global promedio obtenido por los colegios urbanos y rurales; aquí se evidencia que las instituciones ubicadas en zonas rurales obtienen menores niveles de desempeño que los ubicados en la zona urbana.

Figura 13. Desempeño de estudiantes por área de ubicación del colegio



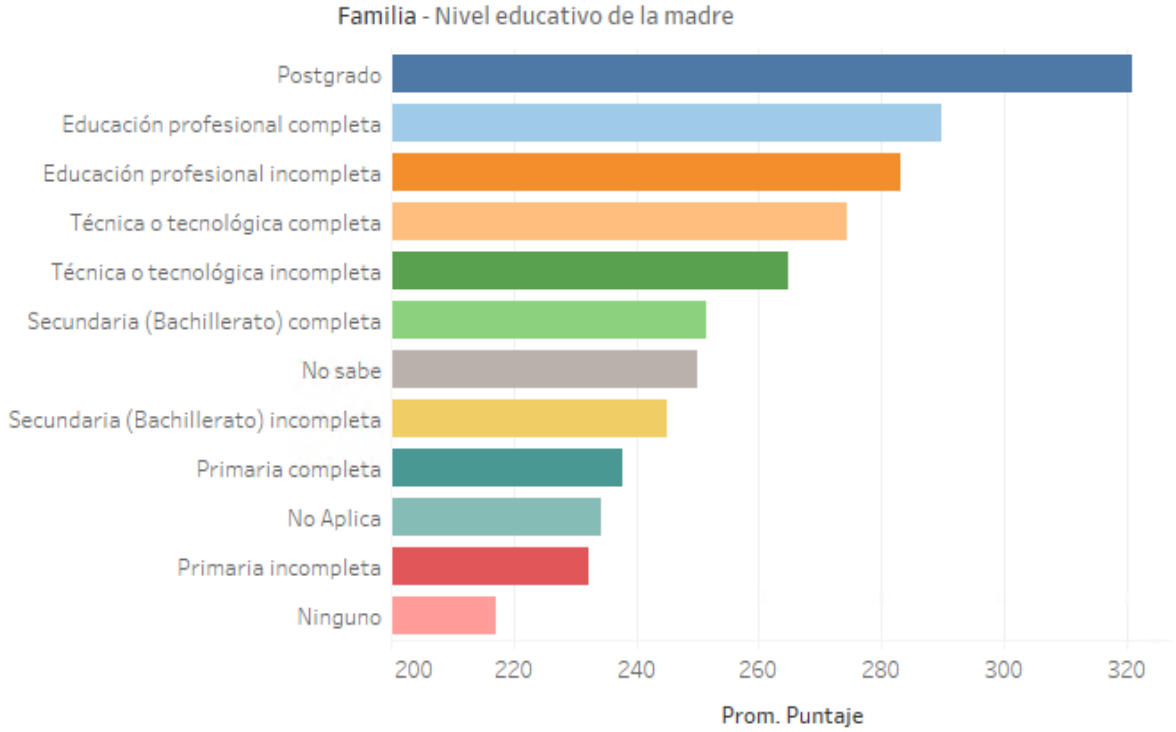
Fuente: Elaboración propia

5. ¿El nivel educativo de los padres influye en el desempeño de los estudiantes?

Para resolver esta inquietud, se analizó el nivel educativo de la madre y el puntaje global obtenido por el evaluado. La figura 14 muestra que el nivel educativo de la madre influye positivamente en los resultados alcanzados; los puntajes promedio

más altos son obtenidos por estudiantes con madres con niveles educativos de posgrado y formación profesional, y los menores niveles de desempeño son obtenidos por estudiantes con madres que no han recibido ningún tipo de educación.

Figura 14. Desempeño de estudiantes por nivel educativo de la madre

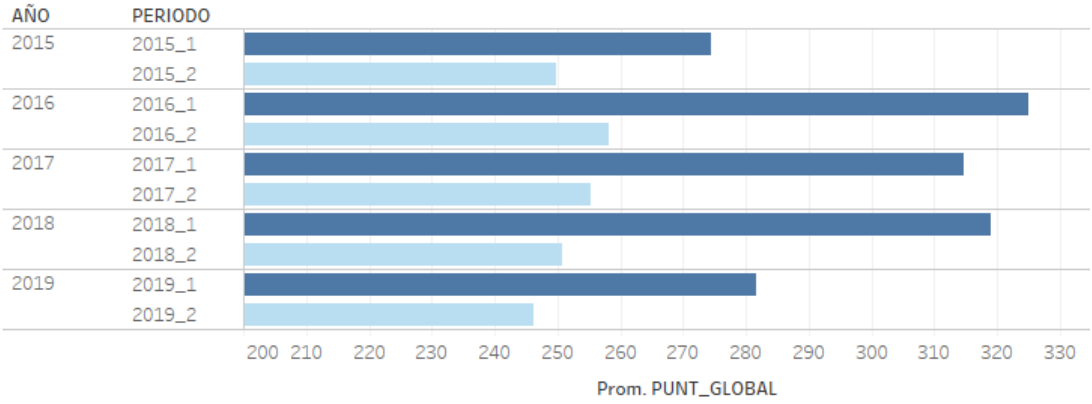


Fuente: Elaboración propia

6. ¿Existe variación entre los resultados obtenidos en el primer y segundo periodo de un año?

Se analizaron los puntajes promedio obtenidos por cada periodo de tiempo evaluados (10 en total). La figura 15 permite determinar que existe una variación constante entre los niveles de desempeño de estudiantes que presentan la prueba en el primer y segundo semestre del año. En todos los años evaluados los mejores puntajes promedio los obtienen los evaluados en el primer semestre, quienes generalmente provienen de colegios no oficiales con calendario B.

Figura 15. Desempeño de estudiantes por semestre del año



Fuente: Elaboración propia

7. ¿Cuál de los grupos étnicos tiene mejor desempeño académico?

Los estudiantes que pertenecen a grupos étnicos suman el 6.028% de la población total evaluada entre los años 2015-2019, y se encuentran distribuidos en 18 etnias: Arhuaco, Cancuamo, Comunidad afrodescendiente, Comunidades Rom (Gitanas), Cubeo, Emberá, Guambiano, Huitoto, Inga, Páez, Palenquero, Pasto, Pijao, Raizal, Sikuani, Tucano, Wayúu, Zenú y un grupo adicional denominado otro grupo étnico minoritario.

Realizando la revisión de la información no se encontró una única etnia que se distinga por sus resultados entre las demás para todos los periodos. La tabla 18 resume las etnias que tuvieron en promedio los mejores desempeños y los resultados menos favorables.

Tabla 18. Etnias con mejores y menores desempeños

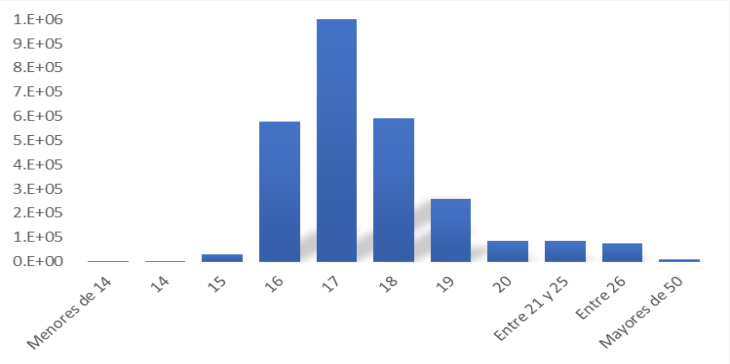
AÑO	MEJOR DESEMPEÑO		MENOR DESEMPEÑO	
	Etnia	Puntaje Promedio	Etnia	Puntaje Promedio
2015	Pasto	251.09	Wayúu	207.96
2016	Comunidades Rom	263.29	Palenquero	206.92
2017	Pasto	258.48	Comunidades Rom	204
2018	Pasto	254.58	Sikuani	195.72
2019	Comunidades Rom	256.36	Sikuani	186.06

Fuente: Elaboración propia

4.2.5. Verificación de la calidad de los datos

Al revisar los valores de fecha de nacimiento, se identificaron años de nacimiento que van desde el año 1 al año 2099; lo que indica claramente valores no posibles. Para el tratamiento de estos datos, se realizó el cálculo de la edad del evaluado teniendo en cuenta su fecha de nacimiento y el año de presentación del examen. La figura 16 presenta la distribución de las edades calculadas, donde se puede identificar que la mayoría de los estudiantes (80%) que presenta el examen tienen entre 16 y 18 años, siendo 17 la edad más común.

Figura 16. Edad de los evaluados



Fuente: Elaboración propia

Debido a que la presentación del examen SABER11 no tiene límite de edad, pero está dirigido a la población bachiller se condicionó los valores posibles de edad entre 13 y 50 años. A los valores que no estén en ese rango se asignó el valor de la media.

La verificación de la calidad de los datos especifica la revisión de datos faltantes (*missings*) en el conjunto de datos, las tablas 9-17 incluyen los valores faltantes para atributos numéricos como categóricos.

ESTU_ETNIA, COLE_BILINGUE y ESTU_RECIBESALARIO son los únicos atributos con faltantes superiores al 5%, sin embargo, anteriormente se mencionó que estos valores no se consideran *missings* sino que toman el valor negativo del parámetro: ninguno, no y no, respectivamente.

4.2.6. Selección de datos

Anteriormente se mencionó la existencia de 60 atributos; sin embargo, algunos de ellos se definirán como atributos No seleccionados ya que no aportan información a la comprensión de la situación académica colombiana:

1. Atributos que representan numeraciones consecutivas como ESTU_CONSECUTIVO y los códigos de: departamentos, municipios, establecimientos educativos, sedes de establecimientos educativos.

2. Atributos que tienen el mismo valor para la totalidad de las instancias como ESTU_ESTUDIANTE que siempre toma el valor ESTUDIANTE.

3. Atributos donde más del 99% de los datos tiene el mismo valor. Como ESTU_NACIONALIDAD, ESTU_PAISRESIDE, ESTU_ESTADOINVESTIGACION.

4. Atributos como tipo de documento, y fecha de nacimiento no se consideran relevantes. Sin embargo, la edad calculada a partir de la fecha de nacimiento y el periodo de presentación del examen si se tienen en cuenta.

Debido a que los archivos de resultados del ICFES se descargan por periodo, no contienen el AÑO; el cual es necesario cuando se maneja un bloque de datos consolidado; por lo tanto, se crea este atributo. La tabla 19 incluye los atributos seleccionados.

Tabla 19. Atributos seleccionados para el análisis de la situación académica.

ATRIBUTOS SELECCIONADOS	
ESTU_GENERO	ESTU_RECIBESALARIO
EDAD	COLE_NOMBRE_ESTABLECIMIENTO
AÑO	COLE_GENERO
PERIODO	COLE_NATURALEZA
ESTU_ETNIA	COLE_CALENDARIO
ESTU_DEPTO_RESIDE	COLE_BILINGUE
ESTU_MCPIO_RESIDE	COLE_CARACTER
FAMI_ESTRATOVIVIENDA	COLE_NOMBRE_SEDE
FAMI_PERSONASHOGAR	COLE_SEDE_PRINCIPAL
FAMI_CUARTOSHOGAR	COLE_AREA_UBICACION
FAMI_EDUCACIONPADRE	COLE_JORNADA

ATRIBUTOS SELECCIONADOS	
FAMI_EDUCACIONMADRE	COLE_MCPIO_UBICACION
FAMI_TRABAJOLABORPADRE	COLE_DEPTO_UBICACION
FAMI_TRABAJOLABORMADRE	ESTU_PRIVADO_LIBERTAD
FAMI_TIENEINTERNET	ESTU_MCPIO_PRESENTACION
FAMI_TIENESERVICIOTV	ESTU_DEPTO_PRESENTACION
FAMI_TIENECOMPUTADOR	PUNT_LECTURA_CRITICA
FAMI_TIENELAVADORA	PUNT_MATEMATICAS
FAMI_TIENEAUTOMOVIL	PUNT_C_NATURALES
FAMI_NUMLIBROS	PUNT_SOCIALES_CIUDADANAS
ESTU_INSE_INDIVIDUAL	PUNT_INGLES
ESTU_NSE_INDIVIDUAL	DESEMP_INGLES
ESTU_TRABAJAACTUALMENTE	PUNT_GLOBAL

Fuente: Elaboración propia

4.2.7. Limpieza de datos

Para la elaboración de la herramienta de visualización, los datos faltantes de un atributo se asumen como nulos y no aportan valor al momento de realizarse un cálculo específico del mismo, por tal razón no se requirió la eliminación de registros con faltantes.

Para el agrupamiento por características similares, se deben alinear los datos a los requerimientos del algoritmo a aplicar, para este caso *k-means*. Este algoritmo es sensible a los datos faltantes. Existen diversas técnicas para eliminación de faltantes, asignación de media, dato más frecuente, interpolación, entre otros. Debido a que los atributos son categóricos se consideró eliminación de faltantes e imputación teniendo en cuenta características comunes. Se consideró eliminación de faltantes para los casos en que no se puede asumir un valor determinado del atributo como: estrato de vivienda, ocupación de los padres, educación de los padres, personas en la familia, cuartos en el hogar, disfrute de computador e internet, posesión de lavadora, automóvil y cantidad de libros. Para el caso de faltantes en los atributos departamento y municipio de residencia del estudiante se asignó el valor de departamento del colegio y municipio de ubicación del colegio.

4.3. SELECCIÓN DE TÉCNICA DE MODELADO

Se ha seleccionado el modelamiento descriptivo, ya que se alinea a los objetivos del proyecto propuesto; el algoritmo elegido como se mencionó anteriormente es el *k-means* (Paredes, 2020) o agrupamiento basado en centroides, el cual es una técnica descriptiva no supervisada que se basa en un número *k* de clústeres a generar basado en la similitud en algunos atributos. Este algoritmo requiere que los atributos a analizar sean numéricos y se encuentren normalizados.

4.4. MODELAMIENTO Y EVALUACIÓN

En este proyecto, el objetivo del modelamiento es realizar un agrupamiento que permita evidenciar la posible relación de los atributos analizados con el desempeño académico. Para la construcción del modelo descriptivo se seleccionaron las variables que contienen información del estudiante, de su familia, nivel socioeconómico e institución educativa.

A partir de los datos de los resultados de las pruebas SABER 11 para los años 2015-2019 que fueron depurados en la fase de entendimiento de los datos, se realizó su preparación para *k-means* normalizando los datos numéricos y convirtiendo las variables categóricas nominales en valores binarios.

- *Normalización de datos numéricos*: datos como puntaje global, puntajes por área, índice socioeconómico del estudiante, entre otros; requieren ser normalizados para obtener valores entre 0 y 1; para esto se requieren identificar los valores mínimo y máximo del atributo en el conjunto de datos y aplicar la ecuación 2.

  $$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad \text{ec. 2}$$

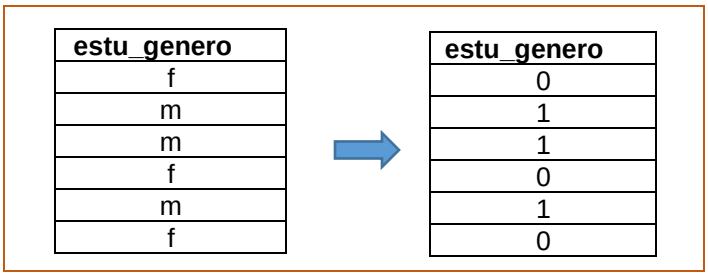
  Ejemplo: Normalizar el puntaje global: 264. Sabiendo que los valores de los puntajes globales en el conjunto de datos se encuentran en el rango [0, 494].

  $$x' = \frac{264 - 0}{494 - 0} = 0.534$$

  De esta forma, al normalizar el puntaje global 264, queda convertido en 0.534.

- *Convertir variables categóricas en numéricas:* Cuando las variables son binarias, es decir que solo pueden tomar uno de dos valores como estu_*genero (f o m)*, o *estu_tieneInternet (s, n)* se transforman los datos y toman el valor de 0 o 1, como se muestra en el ejemplo a continuación en la figura 17.

Figura 17. Conversión de atributo categórico binario en numérico



estu_genero
f
m
m
f
m
f

estu_genero
0
1
1
0
1
0

Fuente: Elaboración propia

Cuando las variables pueden tomar uno de varios valores, como *col_jornada (mañana, tarde, completa, noche, etc.)*, se crean tantas columnas como valores de atributo como se muestra en la figura 18.

Figura 18. Conversión de atributo categórico en numérico

Col_jornada
mañana
tarde
completa
noche
mañana
noche

Col_jornada _mañana	Col_jornada _tarde	Col_jornada _completa	Col_jornada _noche
1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1
1	0	0	0
0	0	0	1

Fuente: Elaboración propia

La herramienta utilizada *KNIME*, permite la normalización y la transformación de variables categóricas a numéricas por medio de los nodos "*Normalize*" y "*One To Many*" respectivamente.

Para determinar el número de grupos (k) en los que se va a clasificar el conjunto de datos para posteriormente realizar el análisis descriptivo, se ejecutó el algoritmo de

agrupamiento con diferentes valores de K y a partir del cálculo del error reportado se determina el número de clústeres adecuado. Es importante aclarar que a mayor cantidad de grupos menor será el error calculado; sin embargo, para este análisis descriptivo a realizar, un número adecuado de clústeres es el que sugiera una disminución de variación de error; y para este caso se ha seleccionado 4. La tabla 20 presenta los errores obtenidos para cada cálculo de k.

Tabla 20. Errores con diferente número de grupos

Número de grupos	Error SSE
2	1.0063E7
3	9296246
4	8937940
5	8603093
6	8471287
7	8346287

Fuente: Elaboración propia

La tabla 21 incluye los puntajes centrales encontrados para cada uno de los clústeres. En esta tabla se pueden diferenciar tres niveles de rendimiento: nivel bajo (clústeres 0 y 1), nivel medio (clúster 2) y nivel alto (clúster 3). Las tablas 22-24 presentan los porcentajes de población de cada clúster asociado con los valores de los atributos relacionados con la familia y del colegio.

Tabla 21.Clústeres: Puntaje

Clúster	Evaluados	PUNTAJE PROMEDIO					
		Global	Matemáticas	Lectura Crítica	Ciencias Naturales	Sociales y Ciudadanas	Inglés
0	27.34%	237.37	46.91	49.63	48.06	45.77	46.08
1	20.06%	239.16	48.03	50.18	47.76	45.90	46.23
2	35.19%	261.70	52.33	54.39	52.17	50.60	51.96
3	17.41%	298.49	60.11	60.34	58.93	58.19	63.35

Fuente: Elaboración propia

Tabla 22. Clústeres: Características familiares

ATRIBUTO		CLUST_0 %	CLUST__1 %	CLUST__2 %	CLUST_3 %
Estrato	Estrato 2	15.31	39.75	57.36	17.89
	Estrato 1	75.71	37.38	18.32	4.19
	Estrato 4	0.87	2.48	2.11	18.80
	Estrato 3	3.81	15.84	20.68	46.09
	Estrato 5	0.40	0.85	0.46	7.86
	Estrato 6	0.27	0.52	0.20	4.74
	Sin estrato	3.62	3.17	0.86	0.43
Personas en el hogar	7 a 8	11.57	9.64	7.11	3.66
	5 a 6	37.45	34.06	32.07	24.89
	3 a 4	39.64	44.21	51.34	63.36
	1a 2	6.77	8.21	6.72	6.72
	9 o más	4.57	3.88	2.76	1.36
Cuartos en el hogar	5 cuartos	2.80	4.38	3.94	4.46
	4 cuartos	9.86	13.83	13.48	16.79
	2 cuartos	47.09	35.65	34.02	22.88
	1 cuarto	7.70	6.36	3.39	1.53
	3 cuartos	31.10	37.30	43.02	52.07
	6 o más	1.44	2.48	2.15	2.26
Internet	Si tiene	11.98	47.00	81.79	96.10
	No tiene	88.02	53.00	18.21	3.90
Computador	No tiene	82.09	52.57	19.96	4.24
	Si tiene	17.91	47.43	80.04	95.76
TV	Si tiene	41.94	70.70	90.62	95.46
	No tiene	58.06	29.30	9.38	4.54
carro	Si tiene	5.07	16.35	15.38	79.30
	No tiene	94.93	83.65	84.62	20.70
Cantidad de libros	11 a 25	23.30	30.93	37.92	24.39
	0 a 10	66.09	47.97	33.57	12.30
	26 a a100	8.60	16.47	22.77	41.96
	Más de 100	2.02	4.64	5.75	21.36

Fuente: Elaboración propia

Tabla 23. Clústeres: Educación de la madre

NIVEL EDUCATIVO	CLUST_0	CLUST_1	CLUST_2	CLUST_3
Técnica o tecnológica incompleta	1.46	2.60	3.61	3.34
Primaria incompleta	28.91	22.16	7.51	1.66
Primaria completa	19.24	15.76	8.69	2.30
secundaria completa	19.77	23.89	37.79	16.02
Secundaria incompleta	17.82	17.37	16.32	5.16
Ninguna	4.26	3.54	0.87	0.27
Profesional completa	2.02	4.24	7.34	37.93
Postgrado	0.18	0.50	0.90	11.41
Técnica o tecnológica completa	4.10	6.73	12.94	14.45
Profesional incompleta	0.68	1.51	2.21	5.33
No sabe/No aplica	1.56	1.71	1.82	2.13

Fuente: Elaboración propia

Tabla 24. Clústeres: Características del colegio

ATRIBUTO	CLUST_0 %	CLUST _1 %	CLUST _2 %	CLUST _3 %
Mixto	98.76	98.89	96.49	88.44
Femenino	1.04	0.76	3.05	8.03
Masculino	0.20	0.35	0.46	3.52
Oficial	92.37	81.24	84.49	17.20
No oficial	7.63	18.76	15.51	82.80
Calendario A	99.67	98.47	98.92	87.17
Calendario B / Otro	0.33	1.53	1.08	12.83
No Bilingüe	98.61	99.02	99.33	94.97
Bilingüe	1.39	0.98	0.67	5.03
Técnico/Académico	34.83	34.19	44.25	8.55
Académico	48.71	53.45	44.99	85.13
Técnico	15.52	11.00	10.20	5.37
No Aplica	0.94	1.35	0.56	0.94
Urbano	71.00	80.25	94.74	93.44
Rural	29.00	19.75	5.26	6.56
Jornada Nocturna	5.95	11.57	3.57	1.53
Jornada Sabatina	6.22	14.02	2.77	1.73
Jornada Mañana	58.04	43.56	58.98	26.55
Jornada Completa	11.54	12.81	8.76	62.90
Jornada Tarde	12.42	11.03	17.76	4.77
Única Jornada	5.84	7.02	8.18	2.52

Fuente: Elaboración propia

A partir de los grupos que se forman para cada conjunto de datos, se identifican las características de los estudiantes que presentaron la prueba en el periodo evaluado.

Clúster 0: Desempeño bajo- NSE bajo

En este clúster se encuentran los evaluados con menores niveles de desempeño a nivel global y en las áreas de matemáticas, lectura crítica sociales y ciudadanas e inglés; es el grupo con mayor porcentaje de evaluados provenientes de colegios oficiales (92%) y de calendario A (99%). Aproximadamente la mitad de evaluados (49%) proviene de colegios de carácter académico. Es el clúster con mayor porcentaje de colegios ubicados en el área rural (29%).

Tiene el mayor porcentaje de estudiantes pertenecientes al estrato 1 (75%). El grupo familiar se distribuye en familias medianas y más del 80% no cuenta con computador ni servicio de internet. Es el grupo que tiene menos libros en casa, 66% de los hogares cuenta con 0 a 10 libros. El nivel educativo de la madre más común es primaria incompleta (29%).

Clúster 1: Desempeño bajo – NSE medio

En este clúster se encuentran los evaluados con menores niveles de desempeño en el área de ciencias naturales; los niveles de desempeño a nivel global y por las demás áreas de conocimiento son levemente superiores a los del clúster 1.

En un 99% está conformado por evaluados de colegio mixto. Es el clúster que agrupa la mayoría de los colegios de jornada nocturna y sabatina (11% y 14% respectivamente). Es el segundo clúster con mayor cantidad de estudiantes provenientes de áreas rurales (19%). La mayoría de su población se distribuye entre los estratos 1 (37%) y 2 (39%); más de la mitad de los evaluados asignados a este grupo no cuenta con computador ni servicio de internet; y de forma similar al clúster 0 cerca del 50% de los hogares pertenecientes al clúster tienen la mínima cantidad de libros en el hogar. El nivel educativo más común de la madre de familia es el bachillerato. Aunque este clúster presenta características de desempeño similares a las del clúster 1 su diferencia se basa en una variación del nivel socioeconómico.

Clúster 2: Desempeño medio

Este grupo contiene puntajes intermedios; con valores superiores a los de los clústeres 0 y 1 pero menores que los del clúster 3. En su mayoría provienen de colegios mixtos (96%) y oficiales (84%) de calendario A. con un 94% perteneciente a colegios urbanos en la jornada de la mañana (59%). El carácter del colegio se distribuye equitativamente entre técnico/académico (44%) y académico (45%)

El 57% de evaluados pertenece al estrato 2 y hay una cantidad representativa que pertenece al estrato 3 (20%). Cerca del 80% cuenta con computador y servicio de internet; se caracteriza por tener más libros en el hogar que los clústeres anteriores (entre 11-25: 38%). La formación educativa más común de la madre es la secundaria completa (38%).

Clúster 3: Desempeño alto

En este clúster se encuentran los puntajes de mejor desempeño en todas las áreas.

En un 82% está conformado por evaluados provenientes de colegios no oficiales, en su mayoría de calendario A. Adicionalmente es el grupo que tiene mayor porcentaje de evaluados de calendario B (12%), de colegios bilingües (5%) y de jornada académica completa (62%). Este grupo es el que tiene un menor porcentaje que pertenece a colegios mixtos (88.8%).

La mayoría pertenece al estrato 3 (46%) y es el que incluye mayor porcentaje de evaluados pertenecientes estratos altos (12% pertenece a los estratos 5 y 6). El grupo familiar se distribuye en hogares pequeños, donde más del 95% cuenta con computador y servicio de internet. Es el grupo con mayor cantidad de libros en el hogar (42% con 26-100 libros). El nivel educativo más común de la madre es profesional completa (35%) y es el grupo con mayor porcentaje de madres con educación de posgrado (11%).

La tabla 25 presenta los porcentajes de población asignados a cada clúster por departamento; como se puede observar cada departamento tiene población evaluada en todos los grupos encontrados.

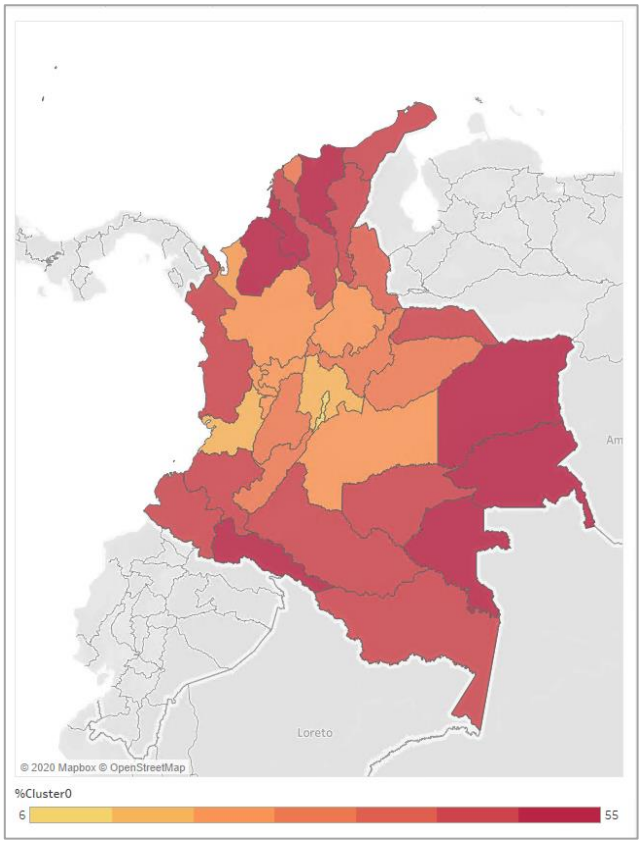
Tabla 25. Clústeres por departamento

Departamento	%Cluster0	%Cluster1	%Cluster2	%Cluster3
AMAZONAS	41.87	24.77	29.02	4.35
ANTIOQUIA	20.43	22.01	42.32	15.24
ARAUCA	42.26	28.67	22.99	6.08
ATLANTICO	33.98	11.63	38.54	15.85
BOGOTA	6.01	16.71	39.81	37.47
BOLIVAR	44.6	13.94	30.62	10.84
BOYACA	29.16	25.7	29.43	15.72
CALDAS	25.67	23.14	37.75	13.44
CAQUETA	41.71	27.62	25.17	5.5
CASANARE	33.16	28.75	31.44	6.65
CAUCA	46.04	26.47	19.59	7.9
CESAR	45.5	14.44	28.92	11.13
CHOCO	45.57	23.39	28.34	2.71
CORDOBA	56.13	16.77	19.41	7.68
CUNDINAMARCA	17.54	24.99	33.75	23.72
GUAINIA	50.95	26.7	21.45	0.89
GUAVIARE	44	29.26	21.2	5.53
HUILA	33.58	29	26.38	11.05
LA GUAJIRA	47.08	17.26	25.55	10.11
MAGDALENA	49.07	15.48	26.86	8.58
META	25.23	23.61	38.72	12.44
NARIÑO	47.68	22.1	21.66	8.56
NORTE DE SANTANDER	34.19	19.66	35.35	10.8
PUTUMAYO	52.57	27.97	17.4	2.06
QUINDIO	23.95	19.1	44.62	12.34
RISARALDA	21.25	21.26	42.24	15.25
SAN ANDRES Y PROVIDENCIA	21.23	19.37	45.13	14.28
SANTANDER	23.58	23.7	37.34	15.38
SUCRE	53.95	14.55	24.73	6.77
TOLIMA	29.9	24.32	34.78	10.99
VALLE DEL CAUCA	18.83	18.23	45.4	17.54
VAUPES	62.42	24.02	13.07	0.49
VICHADA	56.07	25.32	16.21	2.41

Fuente: Elaboración propia

De la tabla anterior se puede observar que los departamentos alejados del centro del país como Vaupés, Córdoba, Vichada, Sucre, Putumayo, Guainía, Magdalena, Nariño, La Guajira, Cauca, Chocó, Cesar, Bolívar, Guaviare, Arauca, Amazonas y Caquetá, concentran su población en el clúster 0 caracterizado por los menores desempeños a nivel nacional y un nivel socioeconómico bajo, estrato 1 en su mayoría, con estudiantes pertenecientes a colegios públicos, educación de los padres en niveles elementales y mayor porcentaje de población rural que los demás grupos. La figura 19 presenta la distribución del clúster 1 a nivel nacional.

Figura 19. Distribución del clúster 0 por departamento

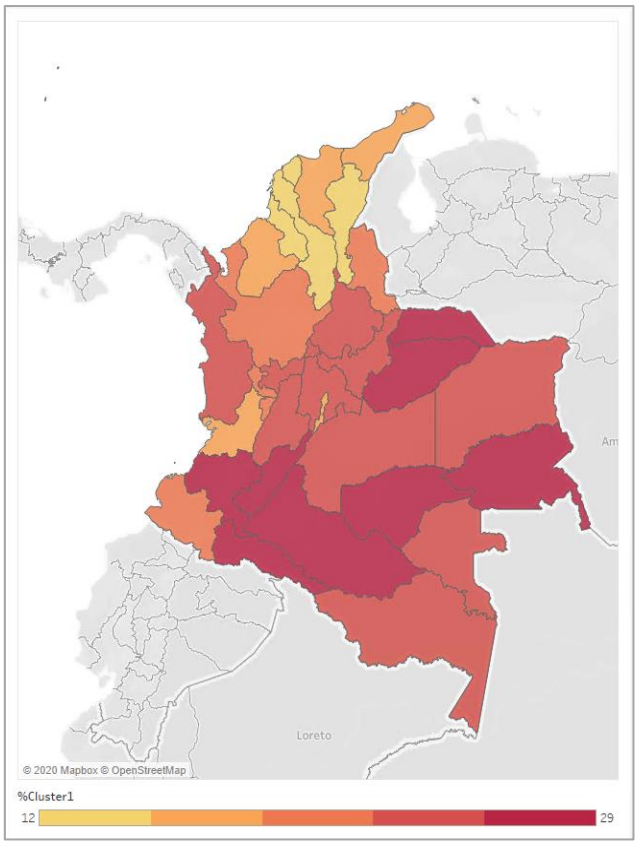


Fuente: Elaboración propia

El clúster 1 caracterizado por obtener puntajes levemente superiores a los del clúster 0, y presentar un incremento sutil en el nivel socioeconómico de los estudiantes, concentra sus evaluados en los departamentos de Guaviare, Huila,

Casanare, Arauca, Putumayo, Caquetá, Guainía y Cauca como se presenta en la figura 20.

Figura 20. Distribución del clúster 1 por departamento



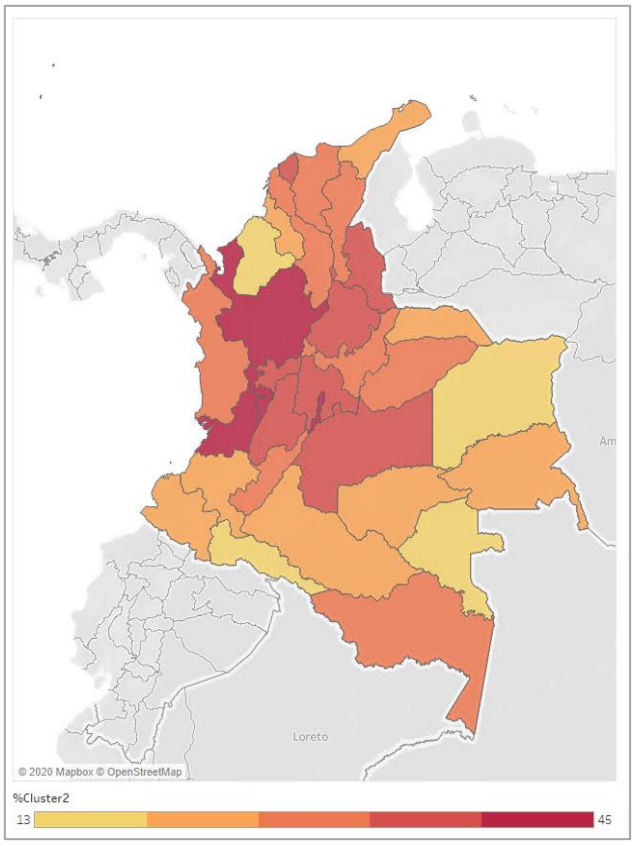
Fuente: Elaboración propia

De acuerdo con los clústeres 0 y 1 se puede observar que los departamentos de Vaupés, Vichada, Putumayo, Guainía, Guaviare, Córdoba, Cauca, Arauca, Nariño, Caquetá, Chocó, Sucre y Amazonas concentran más del 65% de su población en niveles de desempeño bajos.

El clúster 2 incluye la mayor parte de la población evaluada, aunque no es el más popular en la mayoría de los departamentos debido a la variación de la densidad poblacional entre un departamento y otro; como Amazonas y Antioquia que aportan 0.1 y 13% respectivamente de los evaluados totales en este estudio.

El clúster 2 se caracteriza por tener niveles medios de desempeño, mejores a los clústeres 0 y 1 y un nivel socioeconómico que oscila entre los estratos 2 y 3. Bogotá y los departamentos de Valle del Cauca, San Andrés, Quindío, Antioquia, Risaralda, Meta, Atlántico, Caldas, Santander, Norte de Santander, Tolima y Cundinamarca tienen la mayor parte de su población asignada a estos grupos. La figura 21 presenta la distribución de los departamentos.

Figura 21. Distribución del clúster 2 por departamento



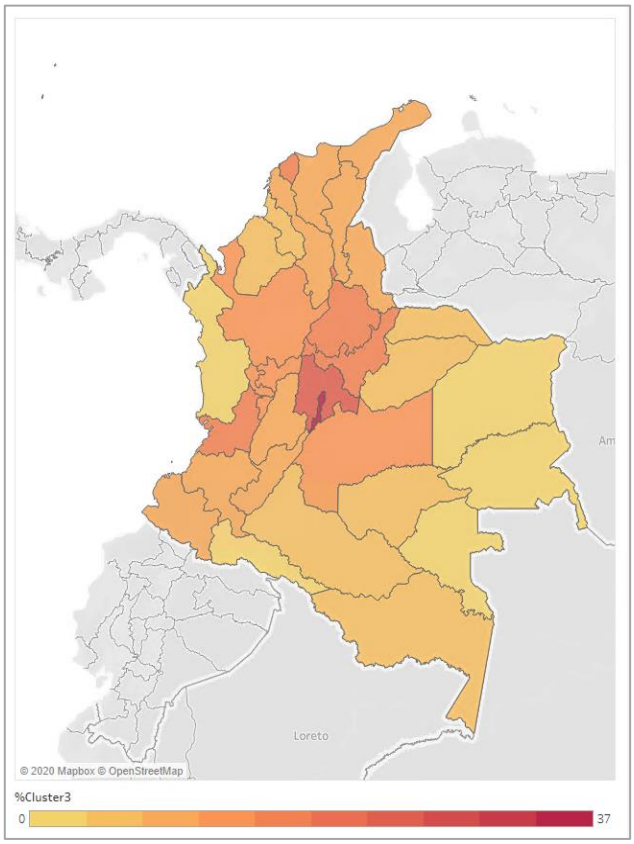
Fuente: Elaboración propia

Finalmente, el clúster 3 caracterizado por obtener los mejores niveles de desempeño y mejor nivel socioeconómico de los evaluados tiene su mayor cantidad de evaluados en Bogotá y en los departamentos donde se encuentran las principales ciudades de país: Cundinamarca, Valle del Cauca, Atlántico, Boyacá y Santander, como se muestra en la figura 22.

Hasta el momento se han mencionado las características generales de los clústeres generados que reflejan la población de estudio. Es importante mencionar que no es posible asignar en su totalidad un departamento a un clúster específico, ya que, debido a la variación de su población, los municipios dentro del departamento pertenecen a diferentes grupos como se muestra en la figura 23, donde se representa la población de Santander asignada al clúster 2 (más común en el dpto.) y se evidencia que solo en las ciudades de Floridablanca, Piedecuesta, Bucaramanga, Girón y Barrancabermeja predomina el grupo en mención.

El listado de población asignada a cada clúster por municipio se encuentra en el anexo externo denominado *Anexo_ClustersMunicipio.xls.*

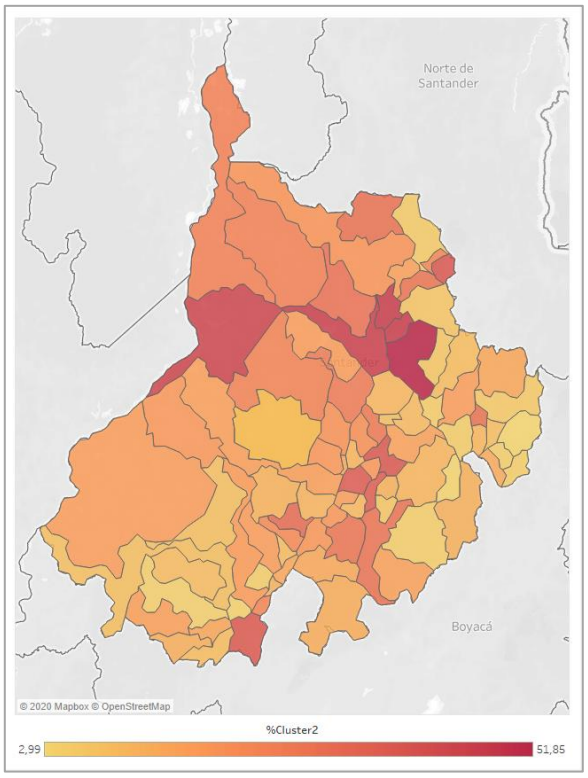
Figura 22. Distribución del clúster 3 por departamento



Fuente: Elaboración propia

El objetivo del agrupamiento realizado en esta etapa era el de encontrar grupos representativos relacionados con el desempeño de los estudiantes, evidenciando que algunas características se relacionan con éste, tales como nivel socioeconómico del estudiante, estrato de la familia, cantidad de libros en el hogar, educación de la madre, acceso a computador e internet y características específicas del establecimiento educativo tales como naturaleza y carácter de la institución.

Figura 23. Distribución del clúster 2 en Santander



Fuente: Elaboración propia

A continuación, se presenta la herramienta de visualización que se desarrolló para el análisis de la situación académica colombiana, donde se verán reflejados los hallazgos encontrados de relación entre el desempeño de los estudiantes y las características mencionadas anteriormente; adicionalmente la herramienta permitirá un análisis geográfico que valida los hallazgos encontrados en el modelamiento sobre el rendimiento de los evaluados en la zona central del país y las zonas alejadas a ésta.

4.5. DESARROLLO DE LA HERRAMIENTA DE VISUALIZACIÓN

Para el desarrollo del mapa interactivo se utilizó la herramienta *Tableau* en su versión Pública. El origen de datos es un archivo denominado *consolidado.csv* exportado del software *KNIME*, el cual contiene la información obtenida en la fase de entendimiento de los datos; los atributos que se analizaron fueron los elegidos en la etapa selección de datos (tabla 19), adicionando los códigos DANE de los departamentos y municipios.

Para realizar el análisis geográfico se implementaron dos mapas de Colombia: uno a nivel departamental y otro a nivel municipal (Ver Anexo B). Debido a que la herramienta gráfica de mapas de *Tableau* no reconoce 250 municipios de Colombia, fue necesario crear un origen de datos adicional con un archivo denominado *lista_mcpios.xlsx,* que incluye los valores de latitud y longitud de cada municipio para relacionarlo a los datos consolidados provenientes de la herramienta de tratamiento de datos *KNIME*.

Dado que el propósito del presente trabajo es proveer una herramienta que permita el análisis de la situación académica del país, se plantearon tres visualizaciones: *mapa general,* el cual cubre la totalidad de la población evaluada; *etnias.* La cual se enfoca en los resultados obtenidos por evaluados pertenecientes a grupos étnicos y *estadísticas* que permite realizar una revisión de la evolución de los resultados por departamento y por municipio y presenta estadísticas de los resultados obtenidos por año.

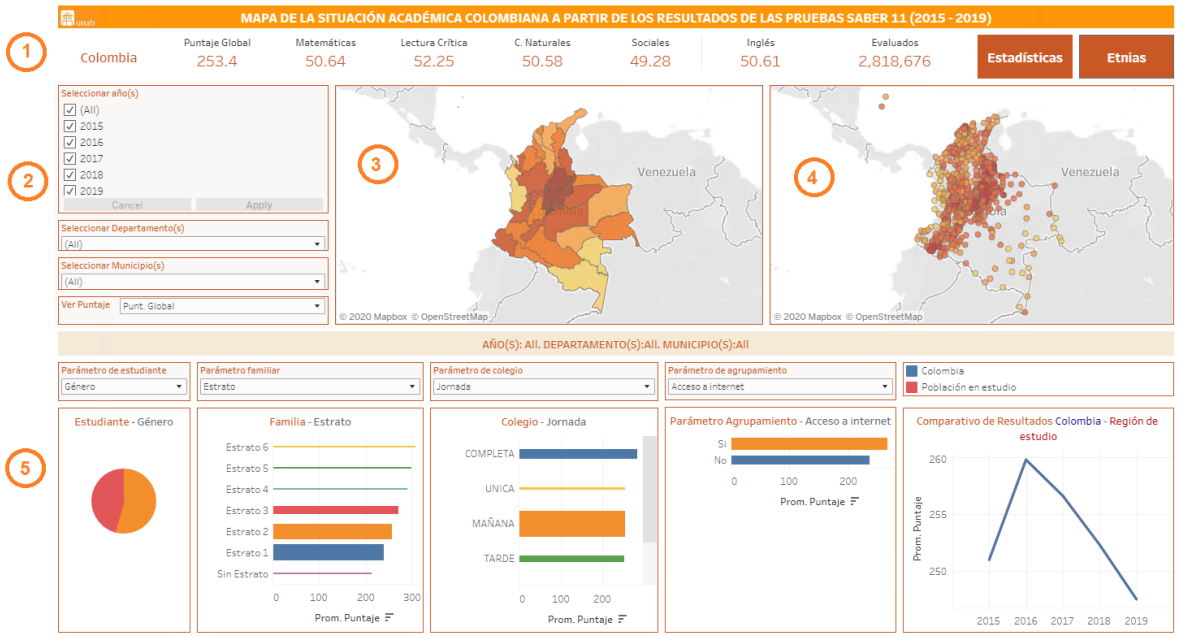
A continuación, se presenta una descripción de la herramienta implementada.

4.5.1. Visualización de mapa población general

Incluye la totalidad de la población evaluada en las pruebas SABER 11 entre los años 2015 y 2019.

La interfaz implementada se presenta en la figura 24 y se compone de cinco secciones.

Figura 24. Interfaz principal mapa interactivo



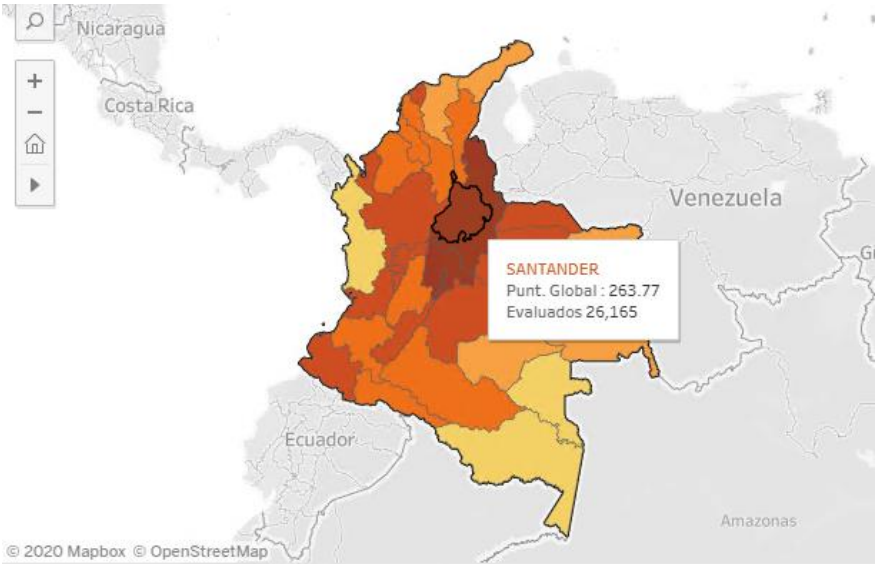
Fuente: Elaboración propia

1. Barra superior: presenta las estadísticas nacionales para el/los años seleccionados y los promedios de los puntajes de cada área de conocimiento evaluada, así como el puntaje global; adicionalmente reporta el número de evaluados en el periodo de tiempo seleccionado.

2. Menú de selección: Ubicado en la parte superior izquierda de la pantalla, permite elegir los periodos a analizar, por defecto están seleccionados todos los años para que muestre las estadísticas acumuladas; sin embargo, es posible seleccionar uno o más años para su análisis. Es necesario siempre oprimir el botón *"Aplicar"* para que los cambios se visualicen. Adicionalmente, este menú permite la selección de departamentos y/o municipios, selección múltiple. De igual forma que en los años, es necesario oprimir el botón *"Aplicar"* para que los cambios se visualicen. Finalmente, se incluye la selección del puntaje que se quiere visualizar y con el que se van a realizar los gráficos: esta opción contiene el puntaje global y los puntajes por área de conocimiento: matemáticas, lectura crítica, ciencias naturales, sociales-y-ciudadanas e inglés.

3. Mapa político de Colombia: Incluye la división en 32 departamentos y el distrito capital. Dependiendo de la selección de departamentos realizada en el panel izquierdo, permite visualizar la información estadística de puntaje promedio (global o por área) obtenido por el departamento y número de estudiantes que presentó la prueba en el periodo de tiempo evaluado. La escala de color está relacionada directamente con el valor del puntaje promedio, los colores claros indican menor desempeño. (figura 25)

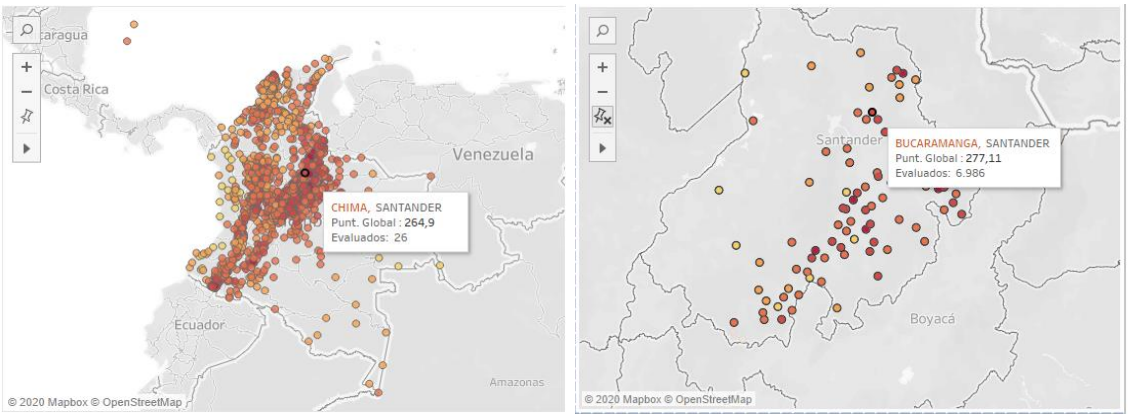
Figura 25. Mapa resultados por departamentos



Fuente: Elaboración propia

4. Mapa de municipios de Colombia: Incluye 1122 municipios de Colombia (incluyendo las áreas no municipalizadas). Dependiendo de la selección de departamentos y municipios realizada en el menú de selección, permite visualizar la información estadística de puntaje promedio (global o por área) obtenido por el municipio y el número de estudiantes que presentó la prueba en el periodo de tiempo evaluado. La ubicación de los municipios se realiza por medio de puntos; al igual que los departamentos maneja la escala de color relacionada con el valor del puntaje promedio como se muestra en la figura 26.

Figura 26. Mapa resultados por municipio



Fuente: Elaboración propia

5. Finalmente, en la parte inferior de la interfaz se encuentran los gráficos que permiten analizar las características del evaluado, su familia y su colegio con los puntajes obtenidos dependiendo de las regiones geográficas seleccionadas. Esta franja se encuentra dividida en seis secciones: título, parámetros del estudiante, parámetros familiares, parámetros del colegio, parámetros de agrupamiento y comparativo de resultados:

   a. Título: Presenta información general de los datos que se están analizando: indica el (los) año(s) en estudio, los departamentos y/o municipios que fueron seleccionados, como se muestra en la figura 27.

   Figura 27. Título de resultados

   AÑO(s): **2019.** DEPARTAMENTO(S): **ANTIOQUIA.** MUNICIPIO(S):**(All).**

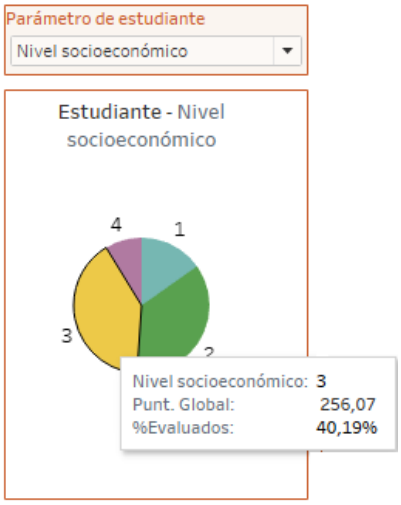
   Fuente: Elaboración propia

   b. Parámetros del estudiante: Por medio de un diagrama de torta permite analizar atributos propios de los estudiantes que pertenecen a la región de estudio como: género, nivel socioeconómico (NSE), si el estudiante trabaja o no y si recibe remuneración.

   Al pasar el cursor sobre la imagen (y sobre todas las imágenes que representen diversos parámetros), se presenta información adicional como

el puntaje promedio y el porcentaje de evaluados pertenecientes a cada atributo evaluado como se muestra en la figura 28.

Figura 28. Gráfico de parámetros de estudiante: nivel socioeconómico



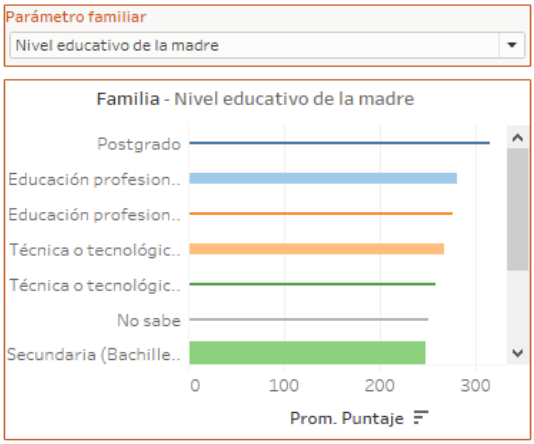
Fuente: Elaboración propia

c. Parámetros familiares: A través de un diagrama de barras permite analizar aspectos familiares de los evaluados de la región de estudio, tales como: estrato familiar, cantidad de personas que viven con el estudiante, ocupación y nivel educativo de los padres, cantidad de libros en casa, y si tiene artículos o servicios como: computador, internet, automóvil, lavadora y TV. La longitud de la barra indica los puntajes promedios por área en estudio y el grosor indica el porcentaje de evaluados que pertenecen a ese valor de atributo como se presenta en la figura 29.

d. Parámetros de colegio: Utilizando un diagrama de barras horizontales permite analizar las características del establecimiento educativo del que proviene el evaluado; incluye los atributos de naturaleza (oficial, privado), carácter o énfasis, jornada, calendario, área de ubicación rural o urbana, género de los estudiantes, si tiene enfoque bilingüe y si es la sede principal del establecimiento. De igual forma que en los parámetros familiares la longitud de la barra indica los puntajes promedio y su grosor indica el

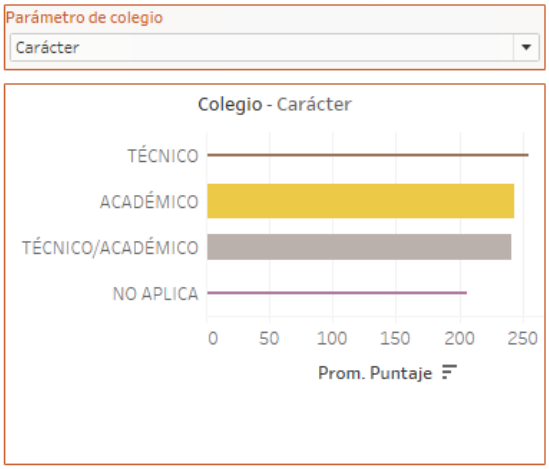
porcentaje de evaluados que pertenece a ese valor de atributo. La figura 30 presenta un ejemplo de un parámetro de colegio.

Figura 29. Gráfico de parámetros familiares: nivel educativo de la madre



Fuente: Elaboración propia

Figura 30. Gráfico de parámetros del colegio: carácter

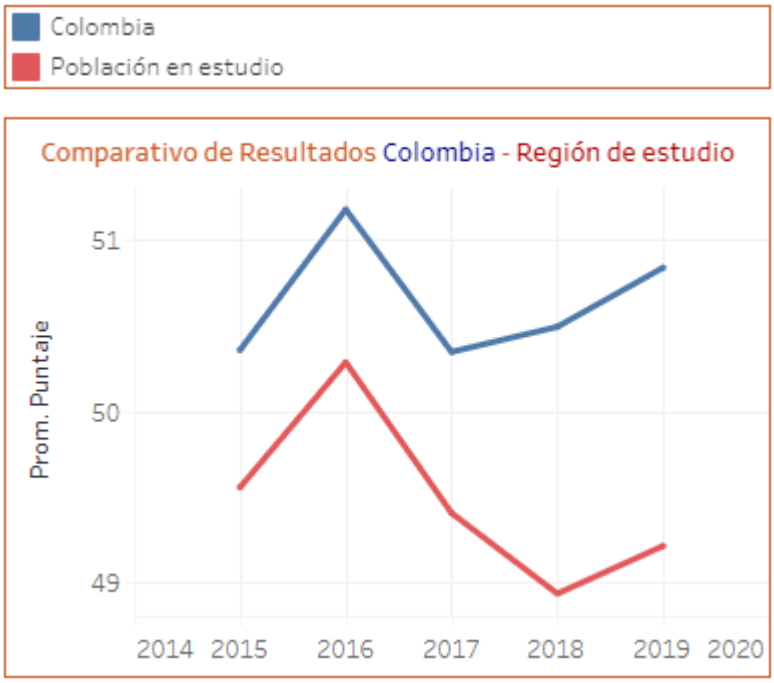


Fuente: Elaboración propia

e. Parámetros de agrupamiento: Utilizando un diagrama de barras horizontales permite analizar las características que se encontraron en la realización de clústeres que influyen directamente en el desempeño del estudiante. Incluye atributos como estrato, acceso a computador e internet, educación de la madre, área de ubicación del colegio, entre otras.

f. Finalmente, por medio de una gráfica de líneas se presenta la evolución de los resultados nacionales y de la región de estudio a través de los años en cada una de las áreas de desempeño evaluadas en la prueba SABER 11, como se muestra en la figura 31.

Figura 31. Gráfico de evolución de resultados por año



Fuente: Elaboración propia

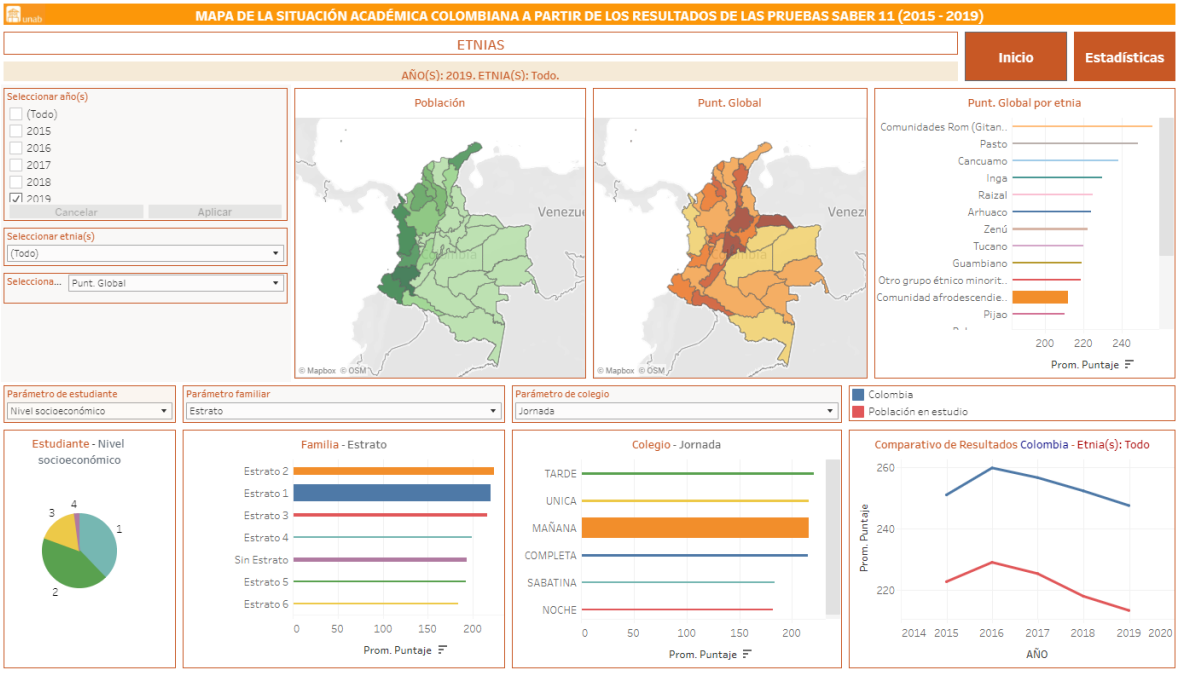
4.5.2. Visualización de mapa por etnias

Incluye la totalidad de la población perteneciente a grupos étnicos evaluada en las pruebas SABER 11 entre los años 2015 y 2019. Los resultados analizados incluyen 18 etnias y un grupo adicional denominado “otros grupos étnicos”.

De forma similar a la población general, la visualización de etnias (ver figura 32) incluye un menú de selección que permite elegir los periodos a analizar, por defecto están seleccionados todos los años para que muestre las estadísticas acumuladas; sin embargo, es posible seleccionar uno o más años para su análisis. Es necesario siempre oprimir el botón “Aplicar” para que los cambios se visualicen.

La interfaz de etnias permite la selección de los grupos étnicos a analizar y del puntaje que se quiere visualizar y con el que se van a calcular las estadísticas: esta opción contiene el puntaje global y los puntajes por área de conocimiento: matemáticas, lectura crítica, ciencias naturales, sociales-y-ciudadanas e inglés.

Figura 32. Interfaz población perteneciente a etnias



Fuente: Elaboración propia

Para la presentación de resultados se implementaron dos mapas políticos de Colombia que presentan información sobre la cantidad de evaluados y los promedios de puntajes obtenidos: el mapa de población (verde) y mapa de puntajes (naranja). El mapa de población indica la cantidad de evaluados pertenecientes a la etnia. Los colores claros indican menor cantidad de evaluados; en el mapa de puntajes, los colores claros indican menores puntajes promedio de desempeño.

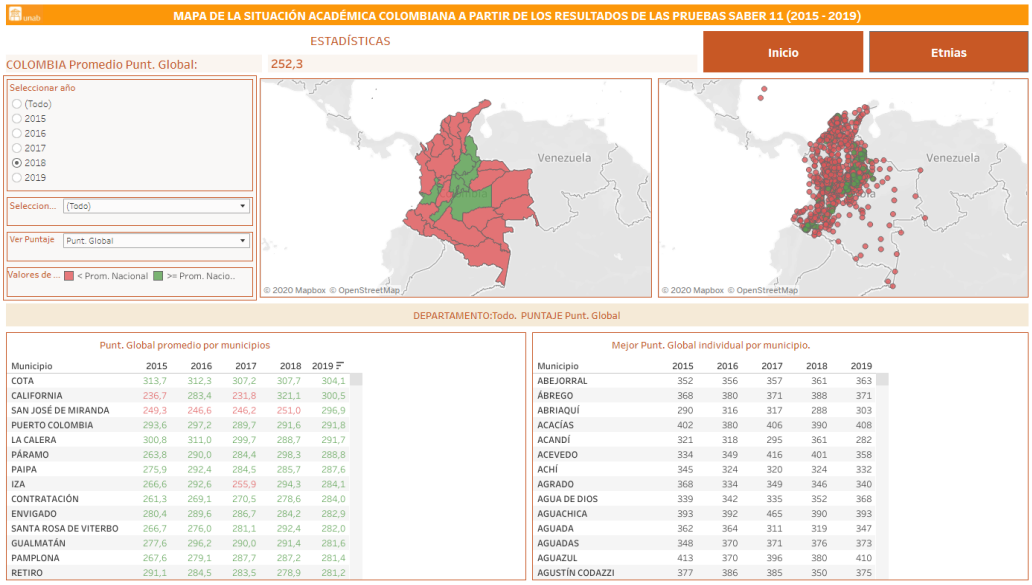
Finalmente, en la parte inferior de la interfaz (ver figura 32) se encuentran las estadísticas de la etnia seleccionada. Esta franja se encuentra dividida en cinco secciones: título, parámetros del estudiante, parámetros familiares, parámetros del

colegio y evolución de resultados nacionales y de las etnias en estudio. Su descripción y funcionalidad es la misma que en la interfaz de población general.

4.5.3. Visualización de estadísticas

Incluye las estadísticas de los puntajes obtenidos por la totalidad de la población. Permite ubicar geográficamente información de los puntajes en las diferentes áreas evaluadas (promedio, mínimo, máximo, desviación) e identifica con verde las zonas que obtuvieron resultados por encima del promedio nacional y en rojo los promedios inferiores, tanto a nivel de departamento como a nivel de municipio (figura 33).

Figura 33. Interfaz estadística



Fuente: Elaboración propia

De forma complementaria en la parte inferior de la pantalla presenta la información tabular de los puntajes promedio para los departamentos y municipios en todos los años analizados, permitiendo identificar los municipios con mejores y menores desempeños en las diferentes áreas evaluadas.

Finalmente, incluye una tabla con los mejores puntajes individuales obtenidos por los estudiantes en los municipios del departamento seleccionado.

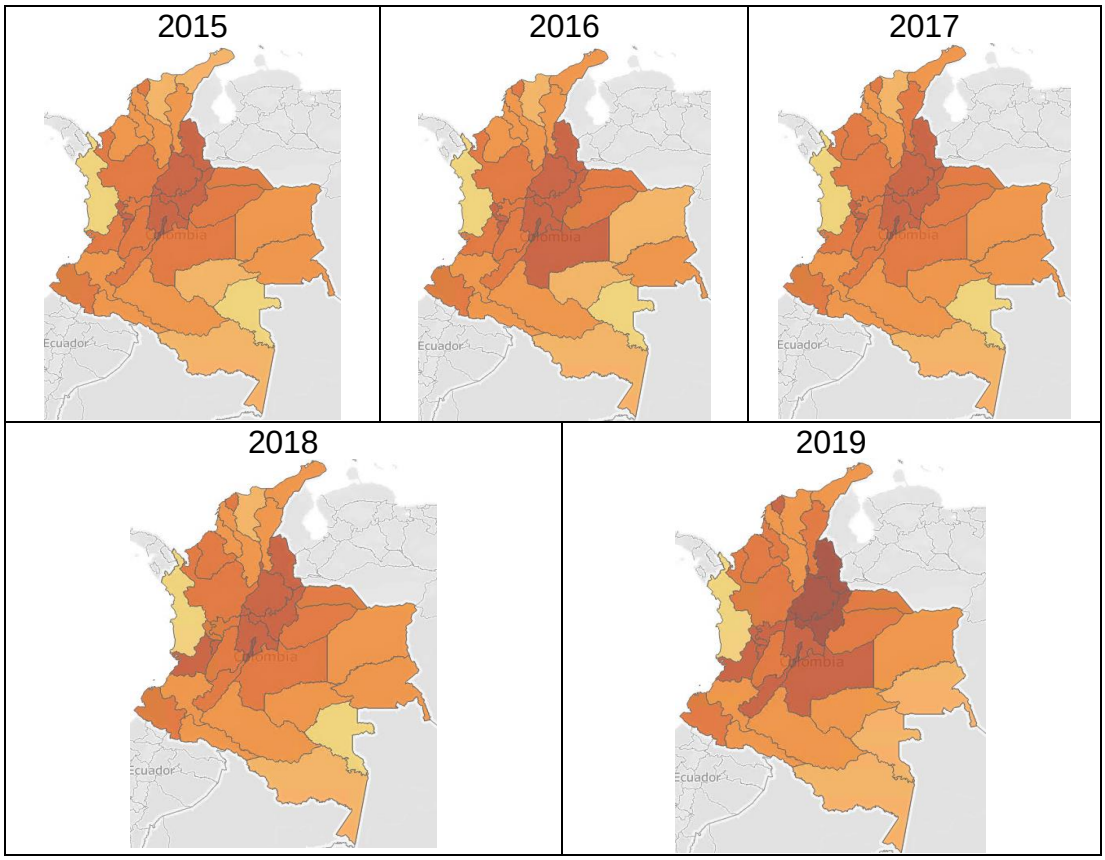
4.6. ANÁLISIS DE LA SITUACIÓN ACADÉMICA COLOMBIANA

Después de realizar el análisis descriptivo de los resultados obtenidos por los estudiantes en las pruebas SABER 11 en el periodo 2015-2019, identificar los grupos con características similares e implementar una herramienta de visualización de la situación académica colombiana se obtienen los siguientes hallazgos.

4.6.1. Población nacional

1. En todos los periodos evaluados se identifica que los mejores resultados se concentran en la zona andina céntrica del país (Bogotá, Cundinamarca, Boyacá, Santander) y los menores desempeños se encuentran en dos departamentos alejados de esta zona: Chocó y Vaupés, como se muestra en la figura 34.

Figura 34. Desempeño por departamento (puntaje global)



Fuente: Elaboración propia

2. La población colombiana que presentó la prueba SABER 11 en el periodo 2015-2019 en su mayoría pertenece a estratos 1 y 2; hace parte de familias medianas entre 3 y 6 personas en el hogar, las madres de los evaluados trabajan en el hogar (45%) y tienen formación de bachiller completa (26%). Los estudiantes tienen acceso a computador (59%) y a servicios de internet (57%). Provienen en su mayoría de colegios oficiales (73%) de carácter académico (54%) de calendario A. El 85% de los evaluados se ubica en área urbana y menos del 2% pertenece a colegios bilingües.

3. Las características socioeconómicas del estudiante y su familia inciden en los resultados de desempeño académico; tanto en los clústeres generados como en el análisis de los mapas se observa que los mejores resultados son obtenidos por evaluados con niveles socioeconómicos mayores de 2, que tienen acceso a computador, servicio de internet y cuentan con varios libros en casa.

4. La educación de los padres también incide directamente en los resultados obtenidos. A nivel nacional y departamental, los menores desempeños se obtienen de evaluados cuyos padres no recibieron educación o solo tuvieron escolaridad básica. Mientras que los padres de los evaluados con resultados más altos tienen formación profesional y de posgrado.

5. Existe una variación significativa entre los resultados obtenidos entre el primer y segundo periodo del año; evidenciando los mejores resultados provenientes de estudiantes que pertenecen a colegios privados, con jornada completa y calendario B.

6. Bogotá y Antioquia tienen la mayor cantidad de evaluados en el periodo de estudio (16.2% y 13.2% respectivamente). Los departamentos de Guainía y Vaupés son los departamentos menor cantidad de evaluados (0.04% y 0.07% respectivamente). La tabla 26 presenta el porcentaje de evaluados por departamento.

Tabla 26. Porcentaje de evaluados por departamento

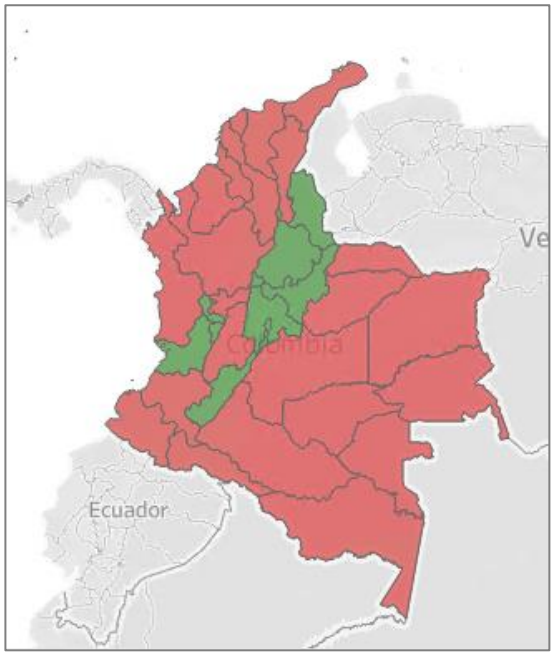
POBLACIÓN EVALUADA POR DEPARTAMENTO	
BOGOTÁ	16,26%
ANTIOQUIA	13,24%
VALLE DEL CAUCA	8,26%
CUNDINAMARCA	6,70%
ATLÁNTICO	5,65%
SANTANDER	4,70%
BOLÍVAR	4,50%
CÓRDOBA	3,46%
NARIÑO	3,07%
TOLIMA	3,03%
BOYACÁ	2,97%
MAGDALENA	2,85%
NORTE DE SANTANDER	2,75%
CAUCA	2,63%
HUILA	2,51%
CESAR	2,29%
META	2,15%
CALDAS	1,93%
SUCRE	1,92%
RISARALDA	1,92%
LA GUAJIRA	1,43%
QUINDÍO	1,25%
CASANARE	1,04%
CHOCÓ	0,78%
CAQUETÁ	0,77%
PUTUMAYO	0,76%
ARAUCA	0,55%
GUAVIARE	0,17%
AMAZONAS	0,15%
SAN ANDRÉS Y PROVIDENCIA	0,12%
VICHADA	0,09%
VAUPÉS	0,07%
GUAINÍA	0,04%

Fuente: Elaboración propia

7. La figura 35 presenta en verde los departamentos que obtienen resultados mayores o iguales a la media y rojo en caso opuesto; aunque solo 9 de los

departamentos de Colombia obtienen puntajes promedio mayores o iguales a la media nacional, éstos representan aproximadamente el 50% de la población total evaluada.

Figura 35. Mapa comparativo con la media nacional



Fuente: Elaboración propia

8. En el periodo 2015-2019 Bogotá y los departamentos de Santander, Boyacá, Cundinamarca, Norte de Santander, Risaralda y Valle del Cauca se han caracterizado por obtener desempeños globales superiores a los del promedio nacional. Los departamentos de Quindío y Huila han ido mejorando sus desempeños, llegando a hacer parte de los puntajes superiores al promedio en los últimos años; en contraste, el departamento de Meta disminuyó sustancialmente su puntaje promedio para el 2019, quedando excluido del grupo. Los demás departamentos (exceptuando Nariño), durante todo el tiempo analizado han presentado un nivel de desempeño inferior al promedio nacional. (ver tabla 27).

Tabla 27. Evolución de resultados por departamento

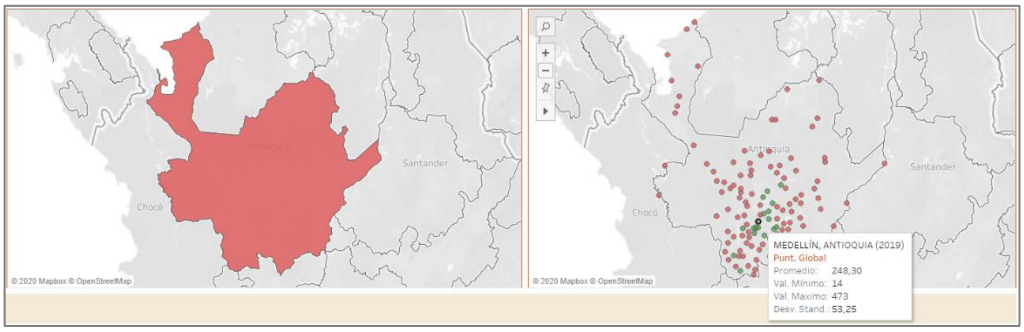
Departamento	2015	2016	2017	2018	2019
BOGOTÁ	270,21	279,59	275,73	272,22	267,27
SANTANDER	259,68	270,17	268,26	266,38	263,77
BOYACÁ	261,11	270,90	268,78	266,56	261,62
CUNDINAMARCA	258,04	267,19	264,01	260,41	255,58
NORTE DE SANTANDER	252,76	264,06	261,29	260,54	254,77
QUINDÍO	251,35	258,32	256,98	255,33	252,29
RISARALDA	255,17	263,39	260,65	256,22	251,41
VALLE DEL CAUCA	251,71	260,97	257,53	253,85	250,51
HUILA	251,16	259,76	253,31	252,79	249,67
META	252,53	262,09	257,54	253,30	247,32
CALDAS	250,88	257,80	256,07	252,12	246,99
CASANARE	249,95	257,93	255,37	251,19	246,98
NARIÑO	250,77	259,27	256,80	251,09	245,38
ATLÁNTICO	245,71	254,47	250,60	247,78	243,14
ANTIOQUIA	247,54	257,12	252,94	247,65	242,19
ARAUCA	246,01	255,07	252,35	247,15	240,65
TOLIMA	244,55	252,15	249,81	245,31	240,62
CESAR	239,55	249,28	247,40	241,56	238,30
PUTUMAYO	239,74	247,59	246,55	240,67	236,45
SUCRE	236,92	248,01	246,70	240,71	235,82
CÓRDOBA	238,59	248,66	246,08	241,20	235,80
CAQUETÁ	239,10	244,62	244,42	238,42	232,64
CAUCA	233,84	242,44	238,63	235,77	231,03
SAN ANDRÉS Y PROVIDEN..	239,37	247,30	252,58	241,39	230,40
BOLÍVAR	232,70	241,31	239,22	233,52	228,57
VICHADA	237,74	238,16	237,69	229,52	225,30
GUAINÍA	242,04	250,00	251,11	235,44	224,68
GUAVIARE	232,49	241,47	235,75	230,49	223,90
MAGDALENA	226,33	233,61	231,33	225,70	221,91
LA GUAJIRA	227,56	232,50	232,21	223,73	218,53
AMAZONAS	220,59	228,08	226,63	216,91	212,96
VAUPÉS	219,64	219,94	221,75	210,90	209,65
CHOCÓ	215,72	219,96	215,39	208,48	200,71

Fuente: Elaboración propia

9. Los departamentos que tienen un puntaje promedio inferior a la media nacional generalmente tienen promedios superiores en sus capitales y en los municipios cercanos a ellas, como se muestra en la figura 36 donde Antioquia, a pesar de

obtener promedios inferiores a nivel departamental, la capital y las ciudades cercanas obtienen niveles de desempeño superiores; esta situación comprueba lo hallado en los clústeres donde, a pesar que un departamento puede tener su mayor porcentaje de población en un clúster determinado; sus municipios pueden pertenecer a otros grupos con niveles de desempeño y características diferentes.

Figura 36. Mapa comparativo con media nacional: Antioquia



Fuente: Elaboración propia

10. De los mejores puntajes obtenidos a nivel de municipio se encontró que en todos los periodos evaluados los municipios de Cota, la Calera y Chía (Cundinamarca) se destacan por obtener los mejores puntajes promedio por municipio del país, como se muestra en la tabla 28.

Tabla 28. Mejores puntajes por municipio

Municipio	Departamento	2015	2016	2017	2018	2019
COTA	CUNDINAMARCA	313,7	312,3	307,2	307,7	304,1
LA CALERA	CUNDINAMARCA	300,8	311,0	299,7	288,7	291,7
SALENTO	QUINDÍO	277,3	289,5	292,4	289,4	280,6
TENJO	CUNDINAMARCA	282,8	283,4	291,0	288,7	278,2
GUALMATÁN	NARIÑO	277,6	296,2	290,0	291,4	281,6
PUERTO COLOMBIA	ATLÁNTICO	293,6	297,2	289,7	291,6	291,8
CHÍA	CUNDINAMARCA	285,4	294,8	288,8	288,9	280,8
PAMPLONA	NORTE DE SANTANDER	267,6	279,1	287,7	287,2	281,4
NOBSA	BOYACÁ	269,8	276,5	287,7	279,7	272,0
FLORIDABLANCA	SANTANDER	276,5	289,7	287,4	282,2	279,6

Fuente: Elaboración propia

11. Los municipios con menores desempeño en los periodos evaluados pertenecen a las regiones amazónica y pacífica del país, predominando municipios rurales de Chocó, Guainía y Amazonas (ver tabla 29).

Tabla 29. Municipios con menores puntajes promedio de desempeño

Municipio	Departamento	2015	2016	2017	2018	2019
TADO	CHOCÓ	206,3	208,9	202,1	201,3	185,0
BAGADÓ	CHOCÓ	205,9	207,6	200,7	184,5	168,4
MAGÜÍ	NARIÑO	204,6	200,2	200,7	190,8	177,6
GUAPI	CAUCA	204,0	208,2	200,2	199,9	189,0
PUERTO NARIÑO	AMAZONAS	198,8	196,8	200,0	193,7	187,2
BAJO BAUDÓ	CHOCÓ	215,2	210,5	199,4	190,3	180,0
CÉRTEGUI	CHOCÓ	209,5	212,9	199,3	196,4	187,7
CERRO DE SAN ANTONIO	MAGDALENA	205,1	210,6	198,9	196,2	186,4
OLAYA HERRERA	NARIÑO	206,4	207,7	198,5	201,2	186,9
PACOA	VAUPÉS		220,4	198,1	189,2	199,0
BOJAYÁ	CHOCÓ	207,0	192,9	198,0	177,4	170,3
TARAPACÁ	AMAZONAS	196,3	214,5	198,0	193,4	200,3
JURADÓ	CHOCÓ	194,1	191,1	197,2	170,2	172,1
ROBERTO PAYÁN	NARIÑO	206,5	199,9	196,2	185,9	179,8
VIGÍA DEL FUERTE	ANTIOQUIA	208,6	210,0	195,3	189,0	190,2
MURINDÓ	ANTIOQUIA	214,3	211,9	195,2	217,6	185,0
ATRATO	CHOCÓ	211,8	208,2	194,8	200,8	199,3
RÍO QUITO	CHOCÓ	203,3	202,5	194,6	191,8	181,2
TIMBIQUÍ	CAUCA	190,4	197,9	194,1	189,4	174,0
RÍO IRÓ	CHOCÓ	194,0	205,8	193,2	188,4	194,3
CARMEN DEL DARIÉN	CHOCÓ	185,1	194,4	192,3	188,7	186,5
MIRITÍ - PARANÁ	AMAZONAS		186,9	192,0	189,2	195,1
LLORÓ	CHOCÓ	191,1	202,4	191,1	180,1	172,9
MEDIO BAUDÓ	CHOCÓ	202,9	196,0	191,0	186,8	188,2
LÓPEZ DE MICAY	CAUCA	193,1	196,1	189,5	179,4	172,6
EL LITORAL DEL SAN JUAN	CHOCÓ	195,5	184,2	185,7	172,2	163,2
ALTO BAUDÓ	CHOCÓ	193,0	188,6	185,6	172,8	157,8
SIPÍ	CHOCÓ	204,6	196,0	178,4	175,4	167,8
PANA PANA	GUAINÍA					177,6
MORICHAL	GUAINÍA				177,4	160,3

Fuente: Elaboración propia

12. La evolución de los puntajes promedio globales y por área de estudio de los municipios de los departamentos de Colombia, se encuentra disponible en la herramienta desarrollada, la cual puede ser accedida ingresando a https://public.tableau.com/profile/andrea1466#!/vizhome/UNAB_ICFES_MAPA_1/Dash_Mapas?publish=yes y seleccionando el botón “Estadísticas”; seleccionando en el menú de la izquierda el año de estudio y el departamento, en el mapa se presentan las estadísticas de cada municipio (mínimo puntaje, máximo puntaje, puntaje promedio, desviación estándar) y en las tablas inferiores los puntajes tabulados de todos los años analizados.

4.6.2. Población perteneciente a grupos étnicos

1. Los estudiantes que pertenecen a grupos étnicos suman el 6.028% de la población total evaluada entre los años 2015-2019, y se encuentran distribuidos en 18 etnias: Arhuaco, Cancuamo, Comunidad afrodescendiente, Comunidades Rom (Gitanas), Cubeo, Emberá, Guambiano, Huitoto, Inga, Páez, Palenquero, Pasto, Pijao, Raizal, Sikuani, Tucano, Wayúu, Zenú y un grupo adicional denominado otro grupo étnico minoritario.

   La tabla 30 presenta los grupos étnicos y la población evaluada, aquí se observa que la comunidad afrodescendiente tiene la mayor población de evaluados de grupos étnicos (56%), y la etnia con menor cantidad de evaluados es la de comunidades ROM (0.04%).

Tabla 30. Evaluados por grupo étnico (2015-2019)

Etnia	% Evaluados de grupos étnicos	Total de evaluados (2015-2019)
Comunidad afrodescendiente	56.07	95.282
Zenú	9.46	16.073
Otro grupo étnico minoritario	9.02	15.331
Wayúu	6.29	10.686
Paez	5.63	9.562
Pasto	4.01	6.814
Emberá	2.91	4.937
Pijao	1.82	3.088
Raizal	1.21	2.051
Inga	0.71	1.213
Sikuani	0.56	946
Guambiano	0.53	893
Cancuamo	0.50	855
Cubeo	0.31	520
Arhuaco	0.31	520
Huitoto	0.27	464
Palenquero	0.23	393
Tucano	0.13	218
Comunidades Rom (Gitanas)	0.04	73

Fuente: Elaboración propia

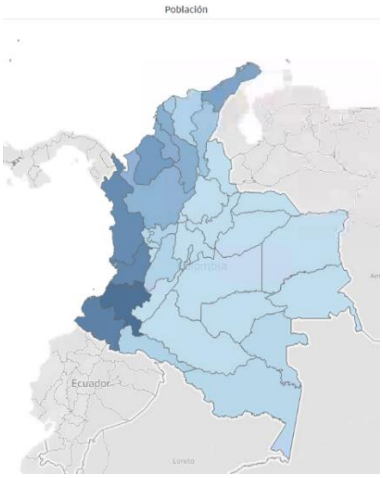
2. Los evaluados se caracterizan por pertenecer al estrato 1, la ocupación de la madre en la mayoría de los casos es el hogar y el padre trabaja de forma independiente o como agricultor/pesquero/jornalero pesquero, la educación de los padres más común es la primaria incompleta. Más del 65% de los evaluados no cuenta con computador ni servicio de internet. El 40% de los estudiantes se concentra en el área rural y el 93% asiste a instituciones oficiales. La figura 37 presenta algunas de estas características.

Figura 37. Características generales grupos étnicos

Fuente: Elaboración propia

3. Los departamentos de la región pacífica: Cauca, Nariño, Valle del Cauca y Chocó concentran la mayor cantidad de evaluados pertenecientes a grupos étnicos, 47% del total del país (figura 38), de ellos, las etnias del departamento de Chocó son las que obtienen menor desempeño.

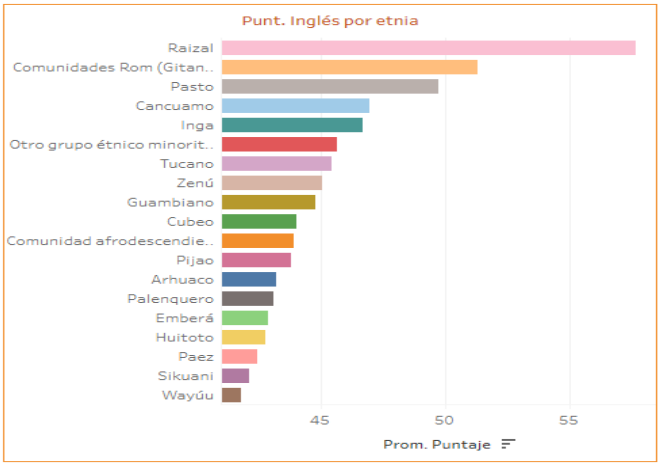
Figura 38. Población étnica



Fuente: Elaboración propia

4. El grupo étnico raizal se distingue por su buen desempeño en inglés, su puntaje promedio es superior al demás etnias (figura 39) y al puntaje promedio nacional obtenido en todos los años evaluados.

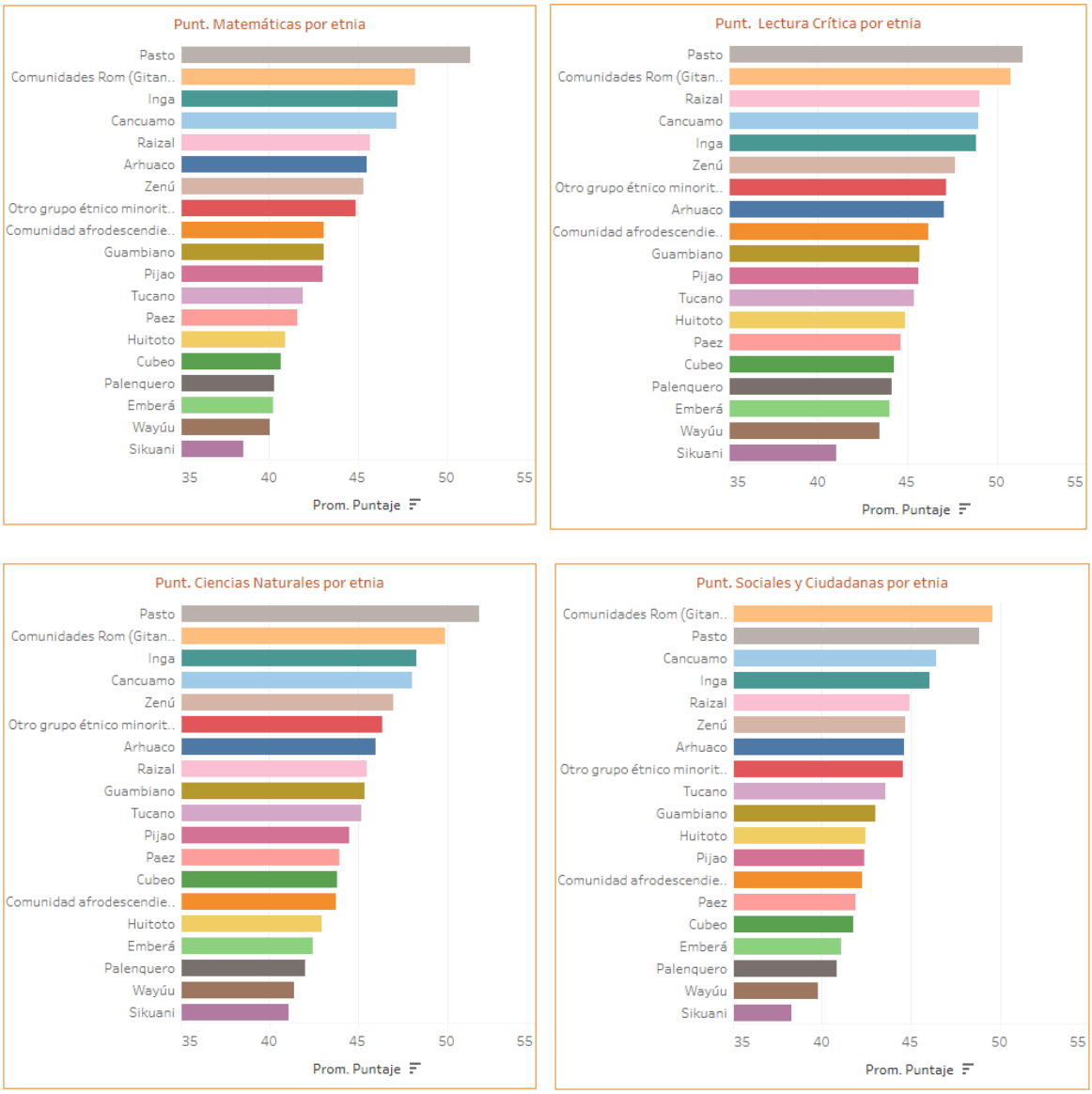
Figura 39. Puntajes promedio de grupos étnicos en inglés



Fuente: Elaboración propia

5. Las etnias pasto y comunidades ROM se identifican por obtener los mejores puntajes promedio en las áreas de conocimiento: matemáticas, lectura crítica, ciencias naturales y sociales-y-ciudadanas; en contraste los grupos étnicos Wayúu y Sikuani, provenientes en su mayoría de La Guajira y Vichada respectivamente, son los que obtienen los puntajes promedio más bajos, como se muestra en la figura 40.

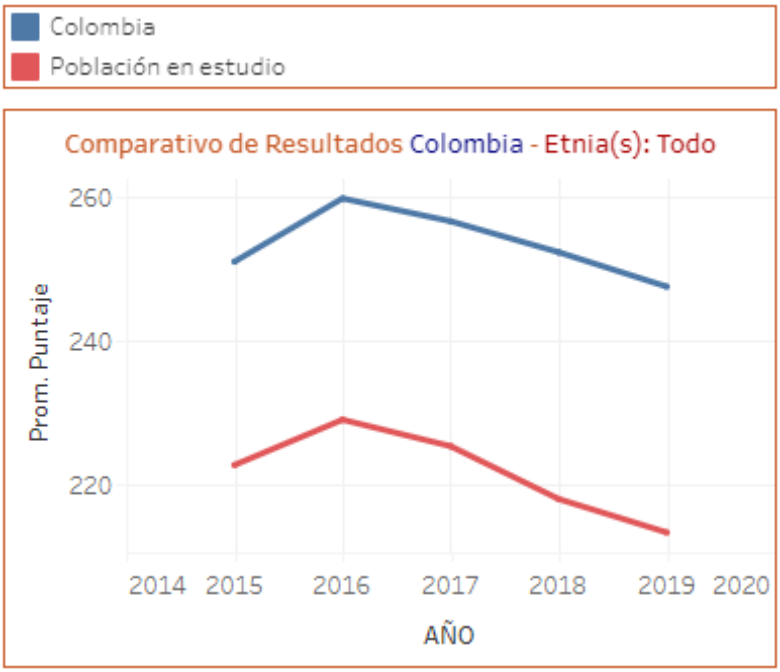
Figura 40. Puntajes promedio de grupos étnicos por área de conocimiento



Fuente: Elaboración propia

6. En todos los años analizados, la población perteneciente a grupos étnicos obtiene puntajes globales promedio inferiores a la media nacional (ver figura 41).

Figura 41. Comparativo de resultados Colombia y grupos étnicos



Fuente: Elaboración propia

4.7. IMPACTO DEL PROYECTO

El desarrollo de este proyecto permitió la aplicación de técnicas de análisis de datos para lograr el desarrollo de una herramienta que suministra información actualizada, organizada y clasificada de la situación académica colombiana a partir del estudio de las bases de datos del ICFES, permitiendo la identificación de los factores que inciden en el rendimiento académico, clasificación de la población estudiantil que presentó la prueba de acuerdo a características comunes y el análisis geográfico del desempeño de los evaluados de la población general y de la población perteneciente a grupos étnicos.

El desarrollo de este proyecto aporta información útil al ministerio de educación nacional, a las secretarías de educación departamentales y/o municipales en el área de fortalecimiento pedagógico, ya que permitirá evaluar el impacto de los procesos pedagógicos, aportando información que apoye la toma decisiones en cuanto a los planes de formación, acompañamiento pedagógico y metas a trazar desde los planes de desarrollo regionales, departamentales y municipales. Adicionalmente, permitirá identificar fortalezas y oportunidades de mejora de los procesos de aprendizaje; aportando conocimiento útil que podrá ser utilizado para el diseño de programas y políticas de mejoramiento de la educación a nivel regional y nacional.

De forma complementaria, el disponer de esta información organizada en una herramienta de visualización que puede ser accedida por el público en general y que permite evaluar el desempeño y las diferentes características socioeconómicas, académicas, personales y familiares de los estudiantes, ofrecerá a diferentes ciudadanos investigadores una herramienta útil para desarrollar propuestas para el mejoramiento de la calidad académica basados en los resultados de los análisis de tendencias y el conocimiento descubierto.

5. CONCLUSIONES

1. Los resultados del análisis realizado muestran que los niveles de desempeño de los estudiantes en las pruebas SABER 11 se encuentran directamente relacionados con las características socioeconómicas de la familia del evaluado y las características del colegio, tanto a nivel municipal, departamental y nacional.
2. La aplicación de técnicas de análisis de datos permite la organización y la comprensión de las diferentes relaciones que pueden existir entre los datos cuando se manejan grandes volúmenes de información.
3. Se ha desarrollado una herramienta de visualización que consolida la información a nivel nacional de resultados de las pruebas SABER 11 de los últimos 5 años, y la presenta de forma organizada y clasificada; permitiendo explorar e identificar los factores que afectan el desempeño académico de los estudiantes y la forma en que han evolucionado los desempeños a nivel nacional, departamental y municipal. La herramienta se encuentra disponible en: https://public.tableau.com/profile/andrea1466#!/vizhome/UNAB_ICFES_MAPA_1/Dash_Mapas?publish=yes
4. El análisis realizado se encuentra limitado a los resultados obtenidos en las pruebas SABER 11, para obtener un análisis más completo de la situación académica colombiana es necesario tomar en cuenta resultados de otras pruebas estandarizadas aplicadas en niveles básicos, medios y profesionales.

6. RECOMENDACIONES Y TRABAJOS FUTUROS

- En un futuro se puede incluir un análisis de los cruces de información de los resultados de las pruebas Saber 11 con los resultados de las pruebas Saber Pro y Saber TyT realizar la trazabilidad de los estudiantes que presentan estas pruebas.
- Una forma de complementar la influencia de la región geográfica en el desempeño de los estudiantes sería incluir fuentes de datos adicionales con información socioeconómica y cultural de los departamentos y municipios para analizar su posible relación con los niveles de desempeño de sus estudiantes.
- Realizar una caracterización del desempeño de los estudiantes de zonas más afectadas por el conflicto armado (ZOMAC).

REFERENCIAS BIBLIOGRÁFICAS

Blanco, V. (2015a). *Análisis del desempeño académico del examen de estado para el ingreso a la educación superior aplicando minería de datos.* Tesis, Universidad Nacional de Colombia.

Blanco, V. (2015b). *Minería de datos en el contexto educativo colombiano: Prueba SABER 11: Mapa Departamento del Cesar*. Obtenido de https://sites.google.com/site/mineriadedatossaber11/

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0 Step-by-step data mining guide.* Obtenido de https://www.the-modeling-agency.com/crisp-dm.pdf

Chávez, D. (2019). Minería de datos para mejorar la educación. *Crónica mx.* Recuperado el 4 de Enero de 2020, de http://www.cronica.com.mx/notas/2019/1111239.html

Consejo Nacional de Política Económica y Social - CONPES. (2018). *Documento CONPES 3920.* Bogotá. Obtenido de https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/3920.pdf

Departamento Nacional de Planeación. (2019). *Bases del Plan Nacional de Desarrollo 2018-2022 – Pacto por Colombia, Pacto por la Equidad.*

European Data Protection Supervisor. (2014). *Report of workshop on Privacy, Consumers, Competition and Big Data.* Bruselas. Obtenido de https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents

Gaviria, C. F., & Posada, H. M. (2019). *Exposición a la contaminación y rendimiento escolar: Evidenciaen los resultados de las pruebas Saber 11 en Colombia.* Universidad de Antioquia, Medellín. Recuperado el 16 de marzo de 2020, de https://www.icfes.gov.co/documents/20143/1645080/1+-+Exposicion+a+la+contaminacion+y+rendimiento+escolar+Evidencia+en+los+resultados+de+las+pruebas+Saber+11+en+Colombia.pdf/2187d0b0-634f-a27d-a453-5cb3d1390778

Gobernación de Amazonas. (2019). *Amazonas avanza, Encuentro regional SABER 2019.* Leticia. Recuperado el Noviembre de 2019, de

http://www.amazonas.gov.co/noticias/amazonas-avanza-encuentro-regional-saber-2019-en-leticia

ICFES. (2015a). *Informe Nacional de Resultados: Colombia en PISA 2015.*

ICFES. (2015b). *Información General Divulgación.* Obtenido de https://www2.icfes.gov.co/en/web/guest/informacion-general-divulgacion

ICFES. (2017a). *Datos abiertos: Visualizaciones de los Mejores Conjuntos*. Obtenido de https://herramientas.datos.gov.co/es/blog/visualizaciones-de-los-mejores-conjuntos

ICFES. (2017b). *Reporteador de históricos*. Recuperado el Diciembre de 2019, de http://www2.icfesinteractivo.gov.co/historicos/

ICFES. (2018a). *Consulta de reportes por entidad territorial*. Recuperado el Enero de 2020, de ttp://www2.icfesinteractivo.gov.co/ReportesSaber359/consultaReporteEntidadTerritorial.jspx

ICFES. (2018a). *Diccionario de variables saber 11 periodo 20142-20182.* Obtenido de https://icfesgovco-my.sharepoint.com/personal/dataicfes_icfes_gov_co/Documents/DataIcfes/4.%20Saber11/2.%20Documentaci%C3%B3n/1.%20Saber11/2.%20Diccionarios%20Saber11/Diccionario%20Saber11%202014-2%20a%202018-2.pdf

ICFES. (2018b). *Documentación del examen SABER 11.*

ICFES. (2018c). *Informes de Resultados de invesigaciones ICFES.* Recuperado el 16 de marzo de 2020, de https://www.icfes.gov.co/web/guest/resultados-de-investigaciones-2017

ICFES. (2019). *Informe nacional de resultados del examen saber 11° 2018.* Bogotá. Recuperado el 15 de marzo de 2020, de https://www.icfes.gov.co/documents/20143/1711757/Informe%20nacional%20resultados%20examen%20saber%2011-%202018.pdf

ICFES. (2019c). *Diccionario de variables SABER 11° periodo 20191-20192.* Obtenido de https://icfesgovco-

my.sharepoint.com/personal/dataicfes_icfes_gov_co/_layouts/15/onedrive.aspx?ct=1589295914045&or=OWA%2DNT&cid=27943174%2D8b52%2Deb60%2Dc2bf%2Dfd8870b02519&originalPath=aHR0cHM6Ly9pY2Zlc2dvdmNvLW15LnNoYXJlcG9pbnQuY29tLzpmOi9nL3BlcnNvbmFsL

ICFES. (2020a). *Bases de datos del ICFES.* Recuperado el 2019, de ftp://200.41.6.169/

ICFES. (2020b). *Icfes resultados SABER*. Obtenido de https://www.icfes.gov.co/resultados-saber-11

ICFES-Ministerio de Educaión Nacional. (2019). ¿Cómo se construye el Índice de Nivel Socioeconómico (INSE) en el contexto de las pruebas Saber? *SABER al detalle*. Obtenido de https://www.icfes.gov.co/documents/20143/1519705/Edicion%204%20-%20boletin%20saber%20al%20detalle.pdf

Jiménez, J. (2018). *Minería de datos educativos: análisis de los factores económicos, sociales y demográficos que influyen en el desempeño de las pruebas saber-pro en estudiantes de ingeniería en Antioquia.* Universidad Pontificia Bolivariana, Medellín.

KNIME A.G. (2020). *Chapter 1. KNIME Analytics Platform: an Overview.* Obtenido de https://www.knime.com/knime-introductory-course/chapter1

Koskela, A. (2016). *Exploring the differences of Finnish students in Pisa 2003 and 2012 using educational datamining.* Tesis, University of Jyväskylä.

Martínez, W., & Turriago, A. (2015). Análisis de distribución geográfica y espacial de los resultados de las Pruebas Saber 11 del Instituto Colombiano para el Fomento de la Educación Superior (ICFES). 2005-2012. *Cuadernos Latinoamericanos de Administración*, 39-49. Recuperado el Diciembre de 2019, de https://www.redalyc.org/articulo.oa?id=409643604005

Medranda, N., Palacios, V., & Moromenacho, T. (2018). Los mapas interactivos, herramientas para la participación ciudadana. *Correspondencias & Análisis*(8), 277-287. doi:https://doi.org/10.24265/cian.2018.n8.14

MinTIC. (2019). *Guía para el uso y aprovechamiento de Datos.* Obtenido de https://gobiernodigital.gov.co/623/articles-9407_guia_datos.pdf

Molina, I., Morales, J., & Rodríguez, S. (2019). *Incidencia de las TIC en el mejoramiento de las pruebas saber 11: un análisis a partir del modelo TPACK.* Bogotá. Recuperado el 16 de marzo de 2020, de https://www.icfes.gov.co/documents/20143/1645080/5+-+Incidencia+de+las+TIC+en+el+mejoramiento+de+las+pruebas+saber+11+un+analisis+a+partir+del+modelo+TPACK.pdf/0f905ba7-2e1e-ef0f-1657-6ce7234745f8

Paredes, D. (2020). *Análisis de Datos y algoritmos de predicción con R.* Quito: Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional.

Rodríguez, C., Maldonado, D., & Barrera, F. (2012). *Calidad de la Educación Básica y Media en Colombia: Diagnóstico y Propuestas.* Universidad de los Andes. Bogotá: Ediciones Uniandes. Recuperado el Diciembre de 2019, de https://www.researchgate.net/publication/248653287_Calidad_de_la_Educacion_Basica_y_Media_en_Colombia_Diagnostico_y_Propuestas

Romero, C., & Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, 40*, 601-618. doi:10.1109/TSMCC.2010.2053532

Roser, M., Nagdy, M., & Ortiz, E. (s.f.). *Our world in data.* Obtenido de Quality of Education: https://ourworldindata.org/quality-of-education

Secretaría de Educación de Bogotá. (2004). *La evaluación: Una estrategia a nivel internacional para el mejoramiento de la calidad educativa.* Obtenido de http://www.eduteka.org/articulos/EvaluacionBogota

Secretaría de Educación de Boyacá. (2019). *Educación para la creatividad y la vida. Histórico de pruebas SAber 3-5-9-11. Años 2012-2018.* Boyacá.

Secretaría de Educación Municipal de Santiago de Cali. (2016). *Colombia-Aplicación, análisis y uso de los resultados de las pruebas SABER.* Santiago de Cali. Recuperado el Noviembre de 2019, de https://planeacion.cali.gov.co/amda/index.php/catalog/24/related_materials

ANEXO A. ANÁLISIS DE PARÁMETROS DE LAS PRUEBAS SABER 11

Tabla A1. Valores del parámetro FAMI_OCUPACIONPADRE

FAMI_ OCUPACIONPADRE 2015-2019	FAMI_ OCUPACIONPADRE 2015-2016	FAMI_ LABORPADRE 2017-2019
Microempresario	Pequeño empresario	Es dueño de un negocio pequeño (tiene pocos empleados o no tiene, por ejemplo, tienda, papelería, etc.
Trabajador por cuenta propia	Trabajador por cuenta propia	Trabaja por cuenta propia (por ejemplo, plomero, electricista)
Empresario o cargo de nivel directivo o gerencial	- Empleado de nivel directivo - Empresario - Empleado con cargo como director o gerente general	Es dueño de un negocio grande, tiene un cargo de nivel directivo o gerencial
Pensionado	Pensionado	Pensionado
Trabaja como profesional o técnico	- Empleado de nivel técnico o profesional - Profesional independiente	Trabaja como profesional (por ejemplo, médico, abogado, ingeniero)
Trabajo de nivel auxiliar o administrativo	Empleado de nivel auxiliar o administrativo	Tiene un trabajo de tipo auxiliar administrativo (por ejemplo, secretario o asistente)
Trabaja en el hogar o no trabaja o no estudia	Hogar	Trabaja en el hogar, no trabaja o estudia
Agricultor -pesquero - jornalero -otra actividad	Otra actividad u ocupación	Es agricultor, pesquero o jornalero
Empleado (vendedor-operario-obrero-conductor-limpieza-mtto-seguridad)	Empleado obrero u operario	- Es vendedor o trabaja en atención al público - Trabaja como personal de limpieza, mantenimiento, seguridad o construcción Es operario de máquinas o conduce vehículos (taxita, chofer)
No sabe - No aplica	No sabe No aplica	

Fuente: Elaboración propia

Tabla A2. Listado de atributos omitidos

ATRIBUTO	2015-1	2015-2	2016-1	2016-2	2017-1	2017-2	2018-1	2018-2	2019-1	2019-2
Información personal										
ESTU_TIENEETNIA		x	x	x	x	x	x	x	x	x
ESTU_LIMITA_MOTRIZ	x	x						x		
ESTU_LIMITA_INVIDENTE	x	x								
ESTU_LIMITA_CONDICIONESPECIAL	x	x								
ESTU_LIMITA_SORDO	x	x								
ESTU_LIMITA_SDOWN	x	x								
ESTU_LIMITA_AUTISMO	x	x								
Información de contacto										
ESTU_AREARESIDE	x	x	x	x						
ESTU_VALORPENSIONCOLEGIO	x	x	x	x						
ESTU_VECESPRESENTOEXAMEN	x	x	x	x						
Información socioeconómica										
FAMI_NUMHERMANOS	x	x	x	x						
FAMI_PISOSHOGAR	x	x	x	x						
FAMI_TELEFONO	x	x	x	x						
FAMI_TIENEHORNO	x	x	x	x						
FAMI_TIENEDVD	x	x	x	x						
FAMI_TIENE_NEVERA										
FAMI_TIENE_CELULAR										
FAMI_TIENEMOTOCICLETA					x	x	x	x	x	x
FAMI_TIENECONSOLAVIDEOJUEGOS					x	x	x	x	x	x
FAMI_INGRESOFMILIARMENSUAL	x	x	x	x						
FAMI_COMELECHEDERIVADOS					x	x	x	x	x	x
FAMI_COMECARNEPESCADOHUEVO					x	x	x	x	x	x
FAMI_COMECEREALFRUTOSLEGUMBRE					x	x	x	x	x	x
FAMI_SITUACIONECONOMICA					x	x	x	x	x	x
ESTU_DEDICACIONLECTURADIARIA					x	x	x	x	x	x
ESTU_DEDICACIONINTERNET					x	x	x	x	x	x
ESTU_HORASSEMANATRABAJA					x	x	x	x	x	x
ESTU_TRABAJAACTUALMENTE	x	x	x	x						
ESTU_RECIBESALARIO	x	x	x	x						
ESTU_TIPOREMUNERACION					x	x	x	x	x	x
Módulo de Antecedentes escolares*	x									

ATRIBUTO	2015-1	2015-2	2016-1	2016-2	2017-1	2017-2	2018-1	2018-2	2019-1	2019-2
Resultados										
PERCENTIL_LECTURA_CRITICA			x	x	x	x	x	x	x	x
DECIL_LECTURA_CRITICA	x	x								
DESEMP_LECTURA_CRITICA			x	x	x	x	x	x	x	x
PERCENTIL_MATEMATICAS			x	x	x	x	x	x	x	x
DESEMP_MATEMATICAS			x	x	x	x	x	x	x	x
DECIL_MATEMATICAS	x	x								
PERCENTIL_C_NATURALES			x	x	x	x	x	x	x	x
DESEMP_C_NATURALES			x	x	x	x	x	x	x	x
DECIL_C_NATURALES	x	x								
PERCENTIL_SOCIALES_CIUDADANAS			x	x	x	x	x	x	x	x
DESEMP_SOCIALES_CIUDADANAS			x	x	x	x	x	x	x	x
PUNT_COMP_CIUDADANA	x	x								
DECIL_COMP_CIUDADANA	x	x								
PUNT_RAZONA_CUANTITATIVO	x	x								
DECIL_RAZONA_CUANTITATIVO	x	x								
PERCENTIL_INGLES			x	x	x	x	x	x	x	x
DECIL_INGLES	x	x								
PERCENTIL_GLOBAL			x	x	x	x	x	x	x	x
ESTU_PUESTO	x	x								
ESTU_NSE_ESTABLECIMIENTO			x	x	x	x	x	x	x	x
ESTU_GENERACION-E								x	x	x
ESTU_PILOPAGA	x	x	x	x		x	x			

* Este módulo solo se realizó para el año 2015. Contiene 44 atributos.

Fuente: Elaboración propia

La tabla A3 presenta el listado de los atributos que se incluyen en el análisis preliminar con los valores que pueden tomar.

Tabla A3. Atributos y valores posibles

ATRIBUTO	VALOR
ESTU_TIPODOCUMENTO	CC – Cédula de ciudadanía CE – Cédula extranjera CR – Certificado registraduría PC – Pasaporte colombiano PE – Pasaporte extranjero RC – Registro civil de nacimiento TI – Tarjeta de identidad <u>Solo en 2015-2018</u> NIP – Número de identificación personal NUI – Número único de identificación PV – Por verificar <u>Solo en 2019</u> CCB – Certificado de cabildo) NES – Número establecido por la SE PEP – Permiso especial de permanencia
ESTU_NACIONALIDAD	Texto
ESTU_GENERO	F - Femenino M - Masculino
ESTU_FECHANACIMIENTO	[DD/MM/AAAA]
PERIODO	20194 (segundo periodo de 2019) 20191 20182 20181 20172 20171 20162 20161 20152 20151
ESTU_CONSECUTIVO	Texto
ESTU_ESTUDIANTE	ESTUDIANTE
ESTU_PAIS_RESIDE	Texto
ESTU_ETNIA	Arhuaco Cancuamo Comunidad afrodescendiente Comunidades Rom (Gitanas) Cubeo Emberá Guambiano Huitoto Inga

ATRIBUTO	VALOR
	Páez Palenquero Pasto Pijao Raizal Sikuani Tucano Wayúu Zenú Otro grupo étnico minoritario Ninguno
ESTU_DEPTO_RESIDE	Texto
ESTU_COD_RESIDE_DEPTO	Numérico
ESTU_MCPIO_RESIDE	Texto
ESTU_COD_RESIDE_MCPIO	Numérico
FAMI_ESTRATOVIVIENDA	Estrato 1 Estrato 2 Estrato 3 Estrato 4 Estrato 5 Estrato 6 Sin Estrato
FAMI_PERSONASHOGAR	1 a 2 3 a 4 5 a 6 7 a 8 9 o más
FAMI_CUARTOSHOGAR	Uno Dos Tres Cuatro Cinco Seis o más
FAMI_EDUCACIONPADRE	Ninguno Primaria incompleta Primaria completa Secundaria (Bachillerato) incompleta Secundaria (Bachillerato) completa Técnica o tecnológica incompleta Técnica o tecnológica completa Educación profesional incompleta Educación profesional completa Postgrado No sabe No aplica

ATRIBUTO	VALOR
FAMI_EDUCACIONMADRE	Ninguno Primaria incompleta Primaria completa Secundaria (Bachillerato) incompleta Secundaria (Bachillerato) completa Técnica o tecnológica incompleta Técnica o tecnológica completa Educación profesional incompleta Educación profesional completa Postgrado No sabe No aplica
FAMI_TRABAJOLABORPADRE	Microempresario Trabajador por cuenta propia Empresario o cargo de nivel directivo o gerencial Pensionado Trabaja como profesional o técnico Trabajo de nivel auxiliar o administrativo Trabaja en el hogar o no trabaja o no estudia Agricultor -pesquero -jornalero -otra actividad Empleado (vendedor-operario-obrero-conductor-limpieza-mtto-seguridad) No sabe - No aplica
FAMI_TRABAJOLABORMADRE	Microempresario Trabajador por cuenta propia Empresario o cargo de nivel directivo o gerencial Pensionado Trabaja como profesional o técnico Trabajo de nivel auxiliar o administrativo Trabaja en el hogar o no trabaja o no estudia Agricultor -pesquero -jornalero -otra actividad Empleado (vendedor-operario-obrero-conductor-limpieza-mtto-seguridad) No sabe - No aplica
FAMI_TIENEINTERNET	No Si
FAMI_TIENESERVICIOTV	No Si

ATRIBUTO	VALOR
FAMI_TIENECOMPUTADOR	No Si
FAMI_TIENELAVADORA	No Si
FAMI_TIENEAUTOMOVIL	No Si
FAMI_NUMLIBROS	0 A 10 LIBROS 11 A 25 LIBROS 26 A 100 LIBROS MÁS DE 100 LIBROS
ESTU_RECIBE SALARIO	SI NO
ESTU_TIPOREMUNERACION	SI NO
COLE_CODIGO_ICFES	Numérico
COLE_COD_DANE_ESTABLECIMIENTO	Numérico
COLE_NOMBRE_ESTABLECIMIENTO	Texto
COLE_GENERO	MIXTO FEMENINO MASCULINO
COLE_NATURALEZA	NO OFICIAL OFICIAL
COLE_CALENDARIO	A B OTRO
COLE_BILINGUE	S N
COLE_CARACTER	ACADÉMICO TÉCNICO TÉCNICO/ACADÉMICO NO APLICA
COLE_COD_DANE_SEDE	Numérico
COLE_NOMBRE_SEDE	Texto
COLE_SEDE_PRINCIPAL	S N
COLE_AREA_UBICACION	RURAL URBANO
COLE_JORNADA	COMPLETA MAÑANA NOCHE SABTINA TARDE UNICA
COLE_COD_MCPIO_UBICACION	Numérico

ATRIBUTO	VALOR
COLE_MCPIO_UBICACION	Texto
COLE_COD_DEPTO_UBICACION	Numérico
COLE_DEPTO_UBICACION	Texto
ESTU_PRIVADO_LIBERTAD	S N
ESTU_COD_MCPIO_PRESENTACION	Numérico
ESTU_MCPIO_PRESENTACION	Texto
ESTU_DEPTO_PRESENTACION	Texto
ESTU_COD_DEPTO_PRESENTACION	Numérico
PUNT_LECTURA_CRITICA	Numérico – Rango [0-100]
PUNT_MATEMATICAS	Numérico – Rango [0-100]
PUNT_C_NATURALES	Numérico – Rango [0-100]
PUNT_SOCIALES_CIUDADANAS	Numérico – Rango [0-100]
PUNT_INGLES	Numérico – Rango [0-100]
DESEMP_INGLES	A- A1 A2 B+ B1
PUNT_GLOBAL	Numérico – Rango [0-500]
ESTU_INSE_INDIVIDUAL	Numérico
ESTU_NSE_INDIVIDUAL	Numérico – Rango [1-4]
ESTU_NSE_ESTABLECIMIENTO	Numérico – Rango [1-4]
ESTU_ESTADOINVESTIGACION	PUBLICAR VALIDEZ OFICINA JURÍDICA <u>Solo en 2015-2018</u> Presente con lectura tardía <u>Solo en 2019</u> No se comprobó identidad del examinado

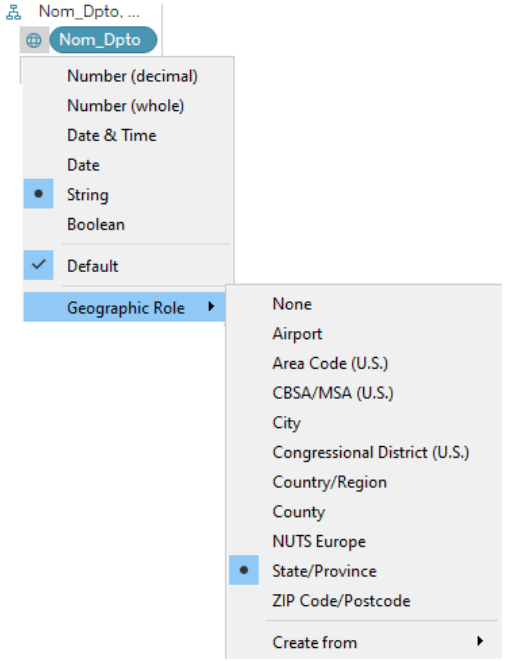
Fuente: Elaboración propia.

ANEXO B. IMPLEMENTACIÓN DE MAPAS INTERACTIVOS EN TABLEAU

1. Implementación de Mapa a nivel de departamento

Los nombres de los departamentos se convirtieron al tipo: función geográfica / estado-provincia como se muestra en la figura B1. Tableau genera las coordenadas de longitud y latitud las cuales se ubican en las columnas y filas del gráfico respectivamente

Figura B1. Tipo de datos de nombre departamento

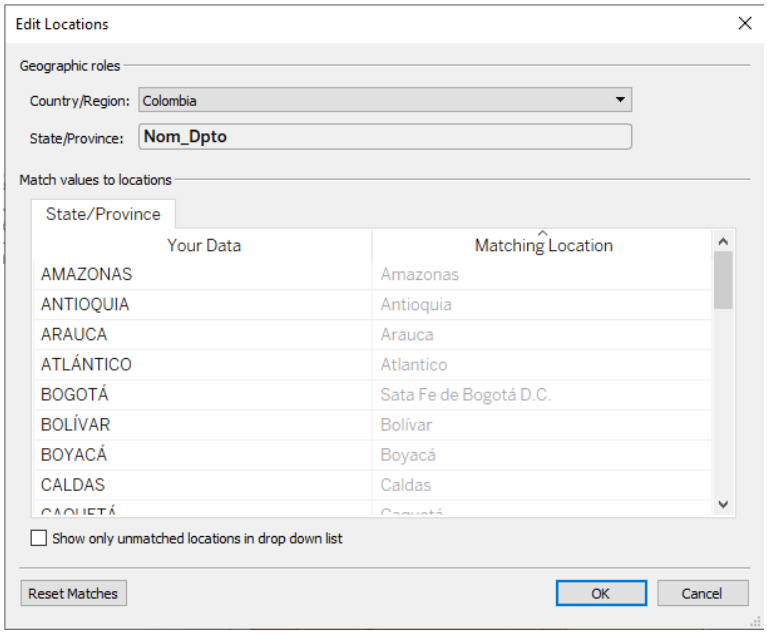


Fuente: Elaboración propia

Por defecto se presenta el mapa de América, para relacionarlo con Colombia, es necesario editar las ubicaciones y elegir a Colombia como el país y la variable con el nombre de los departamentos como el estado o provincia y de esta forma se ubican los departamentos en el mapa de Colombia como lo indica la figura B2

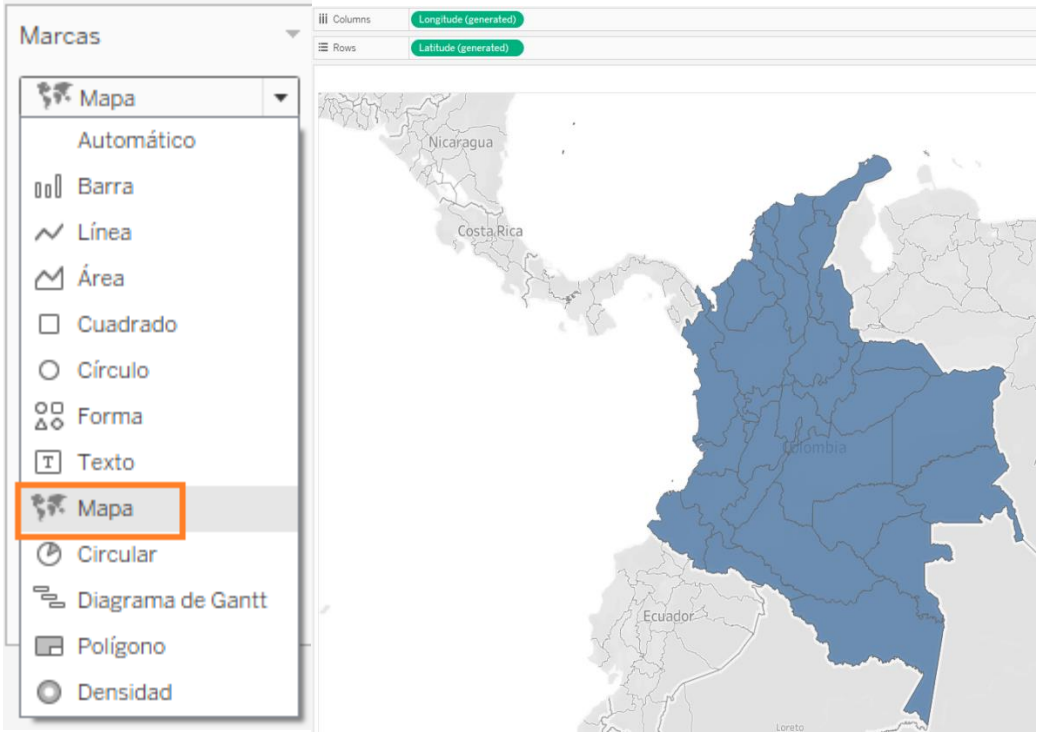
Para que se delinee el contorno del departamento (ver figura B3), es necesario que en las marcas de la imagen se seleccione la opción mapa, de lo contrario solo aparecerá 1 punto por cada departamento, pero no su área correspondiente.

Figura B2. Edición de ubicaciones de mapas en Tableau



Fuente: Elaboración propia

Figura B3. Marcas de imagen de mapas en Tableau

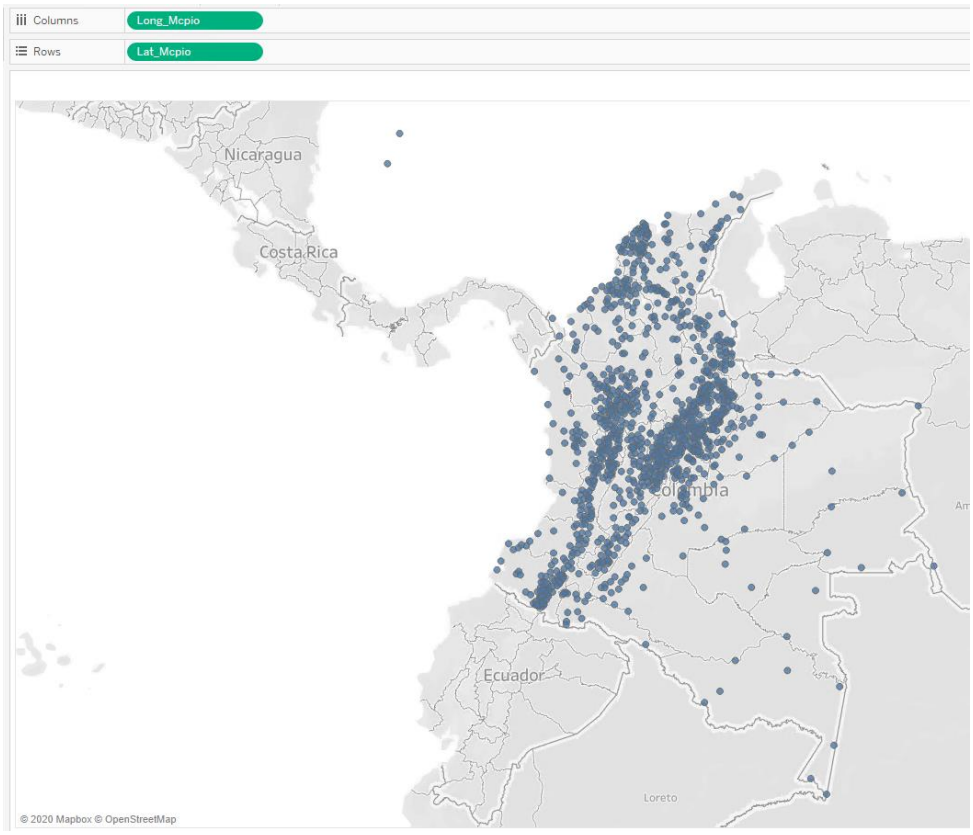


Fuente: Elaboración propia

2. Implementación de Mapa a nivel de municipio

Los nombres de los municipios de Colomba se convirtieron al tipo: función geográfica / ciudad, Tableau genera las coordenadas de latitud y longitud, sin embargo, no identifica 250 municipios y por lo tanto no los ubica en el mapa. Por esta razón es necesario crear un archivo con los valores de longitud y latitud para cada municipio y cruzarlo con la información del proyecto.

Figura B4. Mapa de municipios de Colombia en Tableau



Fuente: Elaboración propia