

**UN MERCADO DE DATOS PARA EL ANÁLISIS MULTIDIMENSIONAL DE
LAS PRUEBAS SABER 11 DE LAS INSTITUCIONES EDUCATIVAS DE LOS
MUNICIPIOS DE LA SUBREGIÓN DE OBANDO DEL DEPARTAMENTO DE
NARIÑO**



Universidad de Nariño

IVAN DAVID VOZMEDIANO BENAVIDES

**UNIVERSIDAD DE NARIÑO
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA DE SISTEMAS
IPIALES
2018**

**UN MERCADO DE DATOS PARA EL ANÁLISIS MULTIDIMENSIONAL DE
LAS PRUEBAS SABER 11 DE LAS INSTITUCIONES EDUCATIVAS DE LOS
MUNICIPIOS DE LA SUBREGIÓN DE OBANDO DEL DEPARTAMENTO DE
NARIÑO.**

IVAN DAVID VOZMEDIANO BENAVIDES

Trabajo de grado presentado como requisito parcial para optar el título de
Ingenieros de Sistemas

**DIRECTOR
PhD. SILVIO RICARDO TIMARÁN PEREIRA**

**UNIVERSIDAD DE NARIÑO
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA DE SISTEMAS
IPIALES
2018**

NOTA DE RESPONSABILIDAD

“La universidad de Nariño no se hace responsable de las opiniones o resultados obtenidos en el presente trabajo y para su publicación priman las normas sobre el derecho de autor”.

Artículo 13, Acuerdo No. 005 del 26 de enero de 2010, emanado por el Honorable Consejo Académico de la Universidad de Nariño.

“Las ideas y conclusiones aportadas en el siguiente trabajo son responsabilidad exclusiva del autor.”

Artículo 1ro del Acuerdo No. 324 de octubre 11 de 1966 emanado del Honorable Consejo Directivo de la Universidad de Nariño.

NOTA DE ACEPTACIÓN:

Firma presidente del jurado

Firma del jurado

Firma del asesor

Firma del co-asesor

AGRADECIMIENTOS

Expreso mi agradecimiento a la Universidad de Nariño por acogerme y formarme como profesional.

Al asesor del trabajo de grado, profesor Ricardo Timaran PhD, por su constante acompañamiento y por permitirme formar parte del grupo de investigación GRIAS para iniciar mi vida investigativa.

A todos los profesores que formaron parte de mi formación, por todas las enseñanzas y conocimientos compartidos.

A todos los compañeros, amigos y demás personas que formaron parte de mi formación académica, por el apoyo y colaboración en esta importante etapa.

DEDICATORIA

Dedico este triunfo a mis padres por el cariño y apoyo incondicional en todos los momentos y etapas de mi vida. Por los consejos recibidos y las enseñanzas aprendidas. Por haberme formado con sus valores y principios para ser un excelente ser humano y profesional.

A mi hermano Wilson Harvey por brindarme día a día momentos de alegría.

A mi abuelo Eduardo Vozmediano por los consejos, el amor y el apoyo incondicional.

A mi abuela Bertha Ligia que, aunque ya no esté en este mundo físicamente, la llevo en el corazón, por sus consejos, enseñanzas, y por ser una dura guerrera de la vida y enseñarme a vivir siempre con alegría.

A mis compañeros Yeison Felicita y Darío Ceballos por el apoyo y los momentos de alegrías y tristezas que compartieron conmigo a lo largo de este proceso.

RESUMEN

La política de la calidad de educación está orientada a que cualquier estudiante independientemente de sus conocimientos particulares cuente con oportunidades para adquirir conocimientos, desarrollar competencias necesarias para vivir en sociedad y ser productivo. A través de los exámenes de estado de calidad de educación media PRUEBA SABER 11°, el Ministerio de Educación Nacional pretende comprobar el grado de desarrollo de las competencias de los estudiantes que cursan el último año escolar (Grado 11°) y de esta forma evaluar la calidad de la educación media.

En este documento se presenta el proceso de diseño y construcción de un mercado de datos que almacena datos históricos de los estudiantes de las Instituciones Educativas de los municipios de la Subregión de Obando del departamento de Nariño que presentaron la Prueba Saber 11° en el período comprendido entre 2012 y 2016, el cual brinda información de calidad para soportar la toma de decisiones acertadas de los directivos de las Instituciones Educativas y secretarías de educación de los municipios en pro del mejoramiento de la calidad de la educación básica y media de esta Subregión.

Palabras clave: Mercado de datos, Prueba Saber 11, Educación media, Subregión de Obando.

ABSTRACT

The education quality policy is oriented to ensure that any student, regardless of their particular knowledge, has opportunities to acquire knowledge, develop skills necessary to live in society and be productive. Through the state exams of quality of secondary education - PROOF KNOW 11, the Ministry of National Education intends to verify the degree of development of the competences of the students that attend the last school year (Grade 11) and in this way evaluate the quality of secondary education.

This document presents the process of design and construction of a data market, which stores historical data of the students of the Educational Institutions of the municipalities of the Obando sub-region of the department of Nariño that presented the SABER 11 TEST in the period covered between 2012 and 2016; provides quality information to support the decision making of the executives of the educational institutions and secretaries of education of the municipalities for the improvement of the quality of basic and secondary education in this Subregion.

Keywords: Data market, Test knowing 11, Middle education, Subregion of Obando.

CONTENIDO

| | Pág. |
|--|-------------|
| INTRODUCCIÓN..... | 13 |
| Prueba saber 11° | 19 |
| Factores Asociados..... | 22 |
| Estructura y alineación | 33 |
| Bodega de datos | 40 |
| Mercado de datos..... | 41 |
| Diseño de un mercado de datos..... | 42 |
| Análisis multidimensional | 44 |
| PostgreSQL..... | 45 |
| PENTAHO | 45 |
| Modelo multidimensional | 46 |
| Operadores OLAP | 48 |
| Subregión de Obando | 49 |
| 1 IMPLEMENTACION | 50 |
| 1.1 Análisis de requerimientos | 50 |
| 1.2 Modelo dimensional | 51 |
| 1.2.1. Diccionario de datos. | 53 |
| 1.3 Diseño del sistema de extracción, transformación y carga (ETL)..... | 57 |
| 1.3.1 Fase de extracción | 57 |
| 1.3.2 Fase de transformación..... | 80 |
| 1.3.3 Fase de carga | 84 |
| 1.4 Implementación del mercado de datos en PostgreSQL | 86 |
| 1.5 Integración del mercado de datos con Pentaho BI | 88 |
| 1.5.1 Cubo de datos | 88 |
| 1.5.2 Operaciones OLAP: | 89 |
| 1.6 Especificación y desarrollo de aplicaciones de BI | 90 |
| 2 DISCUSIÓN Y RESULTADOS..... | 93 |
| 3 CONCLUSIONES..... | 116 |
| 4 RECOMENDACIONES | 117 |
| BIBLIOGRAFÍA..... | 118 |
| ANEXOS..... | 121 |

LISTA DE TABLAS

Pág.

| | |
|---|----|
| Tabla 1. Dimensiones y granularidad del mercado de datos. | 51 |
| Tabla 2. Diccionario de datos de la dimensión estudiante. | 54 |
| Tabla 3. Diccionario de datos de la dimensión instituciones. | 54 |
| Tabla 4. Diccionario de datos de la dimensión lugares. | 54 |
| Tabla 5. Diccionario de datos de la dimensión tiempo. | 55 |
| Tabla 6. Diccionario de datos de la dimensión prueba ciencias naturales. | 55 |
| Tabla 7. Diccionario de datos de la dimensión prueba inglés. | 55 |
| Tabla 8. Diccionario de datos de la dimensión prueba lectura crítica. | 55 |
| Tabla 9. Diccionario de datos de la dimensión prueba matemáticas. | 56 |
| Tabla 10. Diccionario de datos de la dimensión prueba sociales ciudadanas. | 56 |
| Tabla 11. Diccionario de datos de la tabla de hechos. | 56 |
| Tabla 12. Número de registros de cada repositorio. | 57 |
| Tabla 13. Diccionario de datos. | 58 |
| Tabla 14. Atributos de los repositorios fuentes. | 63 |
| Tabla 15. Descripción colores tabla 14. | 63 |
| Tabla 16. Análisis de calidad de datos. | 68 |
| Tabla 17. Operacion OLAP roll up y slice&dice | 89 |
| Tabla 18. Operación olap drill down y slice&dice. | 89 |

LISTA DE FIGURAS

| | Pág. |
|---|--------------------------------------|
| Figura 1. Modelo CIPP..... | 21 |
| Figura 2. Marco de factores asociados..... | 23 |
| Figura 3. Estructura lógica de una bodega de datos..... | 42 |
| Figura 4. Fases de la metodología Kimball..... | 43 |
| Figura 5 .Modelo multidimensional en estrella..... | 47 |
| Figura 6. Modelo multidimensional copo de nieve..... | ¡Error! Marcador no definido. |
| Figura 7. Modelo multidimensional de tipo estrella..... | 53 |
| Figura 8. Configuración de parámetros..... | 84 |
| Figura 9. Proceso carga de datos dimensión estudiante..... | 85 |
| Figura 10. Proceso carga de datos automatizado..... | 86 |
| Figura 11. Modelo Multidimensional tipo estrella en PostgreSQL..... | 87 |
| Figura 12. Parámetros de conexión..... | 88 |
| Figura 13. Creación de cubo en Schema Workbench..... | 89 |
| Figura 15. Pantalla de inicio de la suite de Pentaho..... | 90 |
| Figura 16. Sentencia SQL para diseños de cubos..... | ¡Error! Marcador no definido. |
| Figura 17. Visualización del cubo de datos..... | 90 |
| Figura 18. Gráfico asociado al cubo de datos..... | 92 |
| Figura 19. Número de estudiantes que presentaron la prueba por municipio..... | 93 |
| Figura 20. Número de estudiantes que presentaron la Prueba por año..... | 94 |
| Figura 21. Número de estudiantes por rangos de edades..... | 95 |
| Figura 22. Número de estudiantes que presentaron la prueba de sociales y ciudadanas..... | 96 |
| Figura 23. Promedios de desempeño de la Prueba de sociales y ciudadanas discriminado por rango de edad..... | 97 |
| Figura 24. Promedios de desempeño de la Prueba de ciencias naturales discriminado por rango de edad..... | 98 |
| Figura 25. Promedios de desempeño de la Prueba de sociales y ciudadanas discriminado por rango de edad..... | 99 |
| Figura 26. Promedios de desempeño de la Prueba de sociales y ciudadanas discriminado por municipios..... | 100 |
| Figura 27. Promedios de desempeño de la Prueba de ciencias naturales discriminado por municipios..... | 101 |

| | |
|---|-----|
| Figura 28. Promedios de desempeño de la Prueba de matemáticas discriminado por municipios..... | 102 |
|---|-----|

| | |
|--|-----|
| Figura 29. Número de estudiantes que presentaron la prueba por condición TIC. | 103 |
| Figura 30 Número de estudiantes que presentaron la prueba de sociales y ciudadanas por condición TIC y desempeño de la prueba..... | 104 |
| Figura 31.Promedio de estudiantes que presentaron la prueba de sociales ciudadanas por condición TIC..... | 105 |
| Figura 32.Promedio de estudiantes que presentaron la prueba de ciencias naturales por condición TIC y promedio..... | 105 |
| Figura 33.Promedio de estudiantes que presentaron la prueba de matemáticas por condición TIC y promedio. | 106 |
| Figura 34. Número de estudiantes que presentaron la prueba por género..... | 107 |
| Figura 35.Número de estudiantes que presentaron la prueba por año y género..... | 108 |
| Figura 36. Número de estudiantes que presentaron la prueba de sociales ciudadanas por género. | 109 |
| Figura 37. Número de estudiantes que presentaron la prueba de ciencias naturales por género. | 110 |
| Figura 38.Número de estudiantes que presentaron la prueba de matemáticas por género. | 111 |
| Figura 39.Número de estudiantes que presentaron la prueba de sociales ciudadanas por desempeño y género. | 112 |
| Figura 40.Número de estudiantes que presentaron la prueba de sociales ciudadanas discriminado por desempeño y año. | 113 |
| Figura 41.Número de estudiantes que presentaron la prueba de sociales ciudadanas por desempeño..... | 114 |
| Figura 42.Número de estudiantes que representaron la prueba de estado.... | 115 |

INTRODUCCIÓN

Una de las dificultades presentes a lo largo de la historia del sistema educativo colombiano ha sido la falta de información válida y confiable requerida en el diseño y ejecución de políticas, planes, programas y proyectos; que en conjunto hacen posible la realidad de los propósitos y esfuerzos del gobierno y la sociedad para disponer de una educación de mejor calidad y equidad.

Durante el desarrollo de este proyecto de investigación se construyó un mercado de datos que almacena la información histórica de 18.119 estudiantes de 86 Instituciones Educativas de los municipios de la Subregión de Obando del departamento de Nariño que presentaron la Prueba Saber 11° entre los años 2012 y 2016. La metodología utilizada para diseñar el modelo multidimensional del mercado de datos fue la propuesta por Kimball. El mercado de datos se integró a la herramienta de Inteligencia de Negocios, libre Pentaho; para realizar el análisis multidimensional y la visualización de los resultados. Este mercado de datos brindó información limpia, transformada y de calidad; la cual permitió la generación de indicadores que soporte la toma de decisiones en lo relacionado con la formulación de estrategias para el mejoramiento de la calidad educativa en estas instituciones.

Planteamiento del problema

La función principal de la educación es orientar y apoyar las acciones de mejoramiento de la calidad mediante la obtención, análisis e interpretación de información válida y confiable. Por otra parte, los resultados de pruebas nacionales e internacionales muestran que Colombia posee un sistema educativo con bajos logros académicos de sus estudiantes en cada uno de los niveles de estudio¹

Actualmente el ICFES diseña y aplica las pruebas Saber 3°, Saber 5°, Saber 9° y Saber 11°; con las cuales evalúa la Educación Básica y Media. Estas pruebas evalúan competencias básicas, es decir competencias que se consideran indispensables sin importar la particularidad de las Instituciones Educativas. Su objetivo es contribuir al mejoramiento de la calidad de la educación colombiana mediante la realización de evaluaciones periódicas (censales y muestrales) en las que se valoran las competencias básicas de los estudiantes y se analizan los factores que inciden en sus logros. Los resultados de estas evaluaciones permiten que los establecimientos educativos, las secretarías de educación, el Ministerio de Educación Nacional (MEN) y la sociedad en general conozcan cuáles son las fortalezas y debilidades y a partir de éstas puedan definir planes de mejoramiento en sus respectivos ámbitos de actuación. Su carácter periódico posibilita además valorar cuáles han sido los avances en un determinado lapso de tiempo y establecer el impacto de programas y acciones específicas de mejoramiento².

La Subregión de Obando o Sur es una de las 13 subregiones del departamento de Nariño; la cual está conformada por los municipios de Aldana, Contadero, Córdoba, Cuaspud Carlosama, Cumbal, Funes, Guachucal, Gualmatán, Iles, Ipiales, Potosí, Puerres y Pupiales, que abarcan un total de 4.894 kilómetros cuadrados³.

Uno de los principales problemas que presentan las Instituciones Educativas de los municipios de la Subregión de Obando es el bajo rendimiento que se viene presentando en los últimos años en las Pruebas Saber 11^{o4}. Además, no se

¹ Ver Determinantes del logro académico de los estudiantes de grado 11

² Ver Sistema Nacional de Evaluación Estandarizada

³ Gobernación de Nariño, Ex-Provincia de Obando, 26-11-2012

⁴ ICFES, Consulta de resultados

cuenta con una información histórica de los resultados obtenidos por los estudiantes de esta región en las Pruebas Saber 11° integrados en un mercado de datos que permita obtener resultados confiables y de calidad para la realización de análisis y soporte de decisiones en pro del mejoramiento de la educación en estas Instituciones.

Ni en la Subregión de Obando, ni en otro municipio perteneciente al departamento Nariño se han desarrollado proyectos de investigación que almacenen y analicen los resultados de las Pruebas Saber 11° que presentaron los estudiantes de educación media de las Instituciones Educativas de la región.

En este proyecto de investigación se planteó diseñar un mercado de datos que almacene la información histórica de los estudiantes de las Instituciones Educativas de los municipios de la Subregión de Obando, que presentaron la Prueba Saber 11° entre los años 2012 y 2016. El mercado de datos se integró a la herramienta de Inteligencia de Negocios, libre Pentaho para realizar el análisis multidimensional y la visualización de los resultados. Este mercado de datos brinda información limpia, transformada y de calidad que ayudará a los directivos de las Instituciones Educativas y a la Secretaría de educación de estos municipios a la toma de decisiones en lo relacionado con la formulación de estrategias para el mejoramiento de la calidad educativa en estas Instituciones.

Formulación del problema

¿De qué manera se puede obtener información confiable y de calidad de los resultados de las Pruebas Saber 11° para la realización de análisis y soporte de decisiones, que contribuyan al mejoramiento de la calidad de las Instituciones Educativas de la Subregión de Obando?

Objetivo General. Construir un mercado de datos que almacene datos históricos, limpios y transformados de los resultados de los estudiantes de las Instituciones Educativas de los municipios de la Subregión de Obando del departamento de Nariño que presentaron las Pruebas Saber 11° entre los años 2012 y 2016 que genere información de calidad para soportar la toma de decisiones acertadas de los directivos de las Instituciones Educativas y Secretarías de Educación de los municipios, en pro del mejoramiento de la calidad de la educación básica y media de esta Subregión.

Objetivos Específicos

- Integrar en una base de datos inicial la información de las Pruebas Saber 11°, a partir de las bases de datos del ICFES de los años 2012 al 2016.
- Seleccionar de la base de datos inicial Saber 11°, los atributos socioeconómicos, académicos e institucionales de cada estudiante de las instituciones educativas de la Subregión de Obando y almacenarlos en un repositorio temporal de datos.
- Aplicar técnicas de limpieza y transformación de datos al repositorio temporal de datos de las Pruebas Saber 11°.
- Implementar el proceso de carga de datos al mercado de datos Saber 11° a partir del repositorio temporal de datos.
- Diseñar e implementar el modelo multidimensional en estrella del mercado de datos de las pruebas Saber 11°.
- Integrar el mercado de datos Saber 11° con la herramienta de Inteligencia de Negocios, libre Pentaho.
- Analizar y evaluar los informes y reportes generados por diferentes dimensiones, con la herramienta Pentaho acerca de las Pruebas Saber 11°.

Justificación

Este proyecto de investigación tiene como objetivo diseñar un mercado de datos que almacene la información histórica de los estudiantes de las Instituciones Educativas de los municipios de la Subregión de Obando que presentaron las Pruebas Saber 11° entre los años 2012 y 2016. Este mercado de datos gestionado por una herramienta de Inteligencia de Negocios de software libre como lo es Pentaho, permitirá realizar una vista dimensional de los datos con una máxima efectividad para ganar perspicacia y entender lo necesario para tomar óptimas decisiones que mejoren la calidad de la educación en las Instituciones de esta Subregión.

Por otra parte, un mercado de datos de las Pruebas Saber 11° diseñado bajo un modelo multidimensional y que almacene información limpia y transformada, permitirá la distribución de los resultados del análisis en múltiples formatos y de forma inmediata.

En este sentido, este proyecto se convierte en el primero que se desarrolla en el departamento de Nariño y específicamente en la Subregión de Obando, que busca generar información confiable y de calidad acerca de los resultados de las Pruebas Saber 11° para soportar la toma de decisiones de los directivos de las instituciones educativas y a las secretarías de educación de los municipios

de esta Subregión e incluso al Ministerio de Educación Nacional (MEN), en la formulación de planes que permitan el mejoramiento de la calidad de la educación básica y media de la región.

Contribución de la investigación

Para las instituciones educativas y por ende las secretarías de educación de los municipios de la subregión de Obando, es importante tener información clasificada y de calidad para soportar la toma de decisiones en pro del mejoramiento de la calidad educativa, es así como se realizó el proyecto el cual tuvo como objetivo el diseño y desarrollo de un mercado de datos para el análisis multidimensional, por medio de la metodología presentada por Kimball denominada Ciclo de Vida Dimensional del Negocio, una de las más utilizadas en este tipo de proyectos a través de la herramienta libre Pentaho.

Este trabajo representa un aporte a la comunidad educativa y científica en el área de inteligencia de negocios, específicamente en el descubrimiento de conocimiento en bases de datos, así como a nivel formal y técnico la carga y extracción de datos, consultas SQL y visualización de resultados, representarían una contribución en la investigación básica de la misma área, contando con el respaldo del Grupo de Investigación de Ingeniería de Sistemas de la Universidad de Nariño (GRIAS).

Antecedentes

En la actualidad toda empresa pública o privada necesita herramientas que le permitan depositar mucha confianza en la toma de decisiones sobre los negocios que efectúa. Para la toma de decisiones se necesita analizar hechos y cifras soportados en información confiable, ya que la competitividad en los negocios crece aceleradamente y por ende las decisiones que se deben tomar en la empresa deben ser más rápidas y deben estar basadas en buenos cimientos y para esto se debe manejar y analizar adecuadamente la información que posee la empresa en el menor tiempo posible ya que las demoras pueden causar incluso el cierre del negocio.

En diferentes partes del mundo existen proyectos que utilizan mercados de datos con el fin de obtener información de calidad para soportar las decisiones en diferentes tipos de organizaciones. Entre algunos de ellos se cuenta el caso reportado en Chile con el fin de Diseñar y Construir un mercado de datos para la Mantención de Indicadores de Sostenibilidad de la Industria del Salmón [6]. Como resultado de este proyecto se diseñó un mercado de datos bajo el modelo de tipo constelación en el cual existen 4 tablas de indicadores con algunas dimensiones compartidas que permite ser flexible para diferentes tipos

de consulta relacionados con la industria del Salmón y brinda información confiable, limpia y consolidada para soportar la toma de decisiones.

En México se reporta el proyecto de la Universidad Autónoma de Aguascalientes⁵ que consiste en el diseño de un modelo multidimensional de mercado de datos del área de capacitación para el INEGI, con el propósito de apoyar la toma de decisiones estratégicas y tácticas de dicha área relacionadas con la evaluación de la efectividad y desempeño del programa de capacitación en el instituto que permitan tomar los correctivos necesarios para alcanzar las metas que se proponen los directivos de esta área.

En Colombia se reporta el proyecto de construcción de un mercado de datos para apoyo a las estrategias de permanencia estudiantil realizadas por el área de bienestar institucional en la Escuela Colombiana de Carreras Industriales (ECCI)⁶. Se basa en el análisis, diseño, construcción e implantación de un mercado de datos, apoyado en la metodología de Kimball, que reúne en un solo sitio los diferentes procesos asociados a la identificación, seguimiento y acompañamiento de la población estudiantil con posibles vulnerabilidades de deserción, para lo cual se han identificado todos los posibles escenarios delimitados en el alcance del proyecto. De este modo se detalla en esta fase inicial del proyecto lo evidenciado a partir del análisis y el levantamiento de información con los líderes de los procesos directamente implicados, con el objetivo principal de definir los requerimientos a nivel funcional y técnico permitiendo así realizar los diseños tanto de las fuentes de información no sistematizadas y la definición general de la solución del mercado de datos.

En la Universidad de Manizales se reporta el proyecto denominado Inteligencia de Negocios aplicada a los procesos de autoevaluación de la Universidad de Manizales⁷, en el cual, uno de sus resultados es la construcción de un mercado de datos que contienen información relacionada con los programas, estudiantes, docentes y demás componentes de la institución que deben ser identificados para enfrentar los procesos de autoevaluación y de esta manera seguir con los parámetros propuestos que se dirigen a la Acreditación Institucional.

⁵ Diseño de un modelo multidimensional de data mart del área de capacitación para el INEGI

⁶ DATAMART para apoyo a las estrategias de permanencia estudiantil en la Escuela Colombiana de carreras Industriales

⁷ Inteligencia de Negocios aplicada a los procesos de autoevaluación de la Universidad de Manizales

A nivel regional, en la Universidad de Nariño se desarrolló el proyecto denominado "Diseño de un Data Mart de información académica para la Universidad de Nariño"⁸, cuyo objetivo fue diseñar un mercado de datos con la información histórica de la Oficina de Registro Académico de la Universidad de Nariño, como el primer paso para contar, en un futuro, con un medio eficaz que permita a los directivos de la Universidad de Nariño, soportar la toma de decisiones relacionadas con todos los aspectos académicos, de una manera acertada. Este proyecto quedó en la etapa de diseño y no se implementó.

Fundamentos teóricos

Prueba saber 11°

La Empresa estatal de carácter social ICFES (Instituto Colombiano De Fomento a la Educación Superior), es la encargada de evaluar la educación en todos sus niveles y hacer investigaciones sobre los factores que inciden en la calidad educativa, con la finalidad de ofrecer información para mejorar la calidad de la educación⁹. La Prueba Saber 11° que aplica el Instituto Colombiano para la Evaluación de la Educación media, es un instrumento estandarizado para la evaluación externa y hace parte de los instrumentos que conforman el Sistema Nacional de Evaluación¹⁰.

La Prueba Saber 11° está dirigida a las personas que terminan el nivel de educación media o quienes deseen acreditar que han obtenido los conocimientos esperados de quienes terminaron dicho nivel¹¹; también cabe resaltar que esta prueba es obligatoria para poder ingresar a la educación Superior.

La Prueba Saber 11° es una herramienta que permite evaluar la calidad de la educación en todos los establecimientos educativos del país tanto públicos como privados. El propósito de aplicar estas pruebas es monitorear el rendimiento académico, entendido como el desarrollo de competencias básicas

⁸ Diseño de un Data Mart de información académica para la Universidad de Nariño

⁹ Artículo 12. Ministerio de Educación Nacional. [LEY 1324 DE 2009]

¹⁰ Sistema Nacional de Evaluación Estandarizada

¹¹ Artículo 7. Ministerio de Educación Nacional. [LEY 1324 DE 2009]

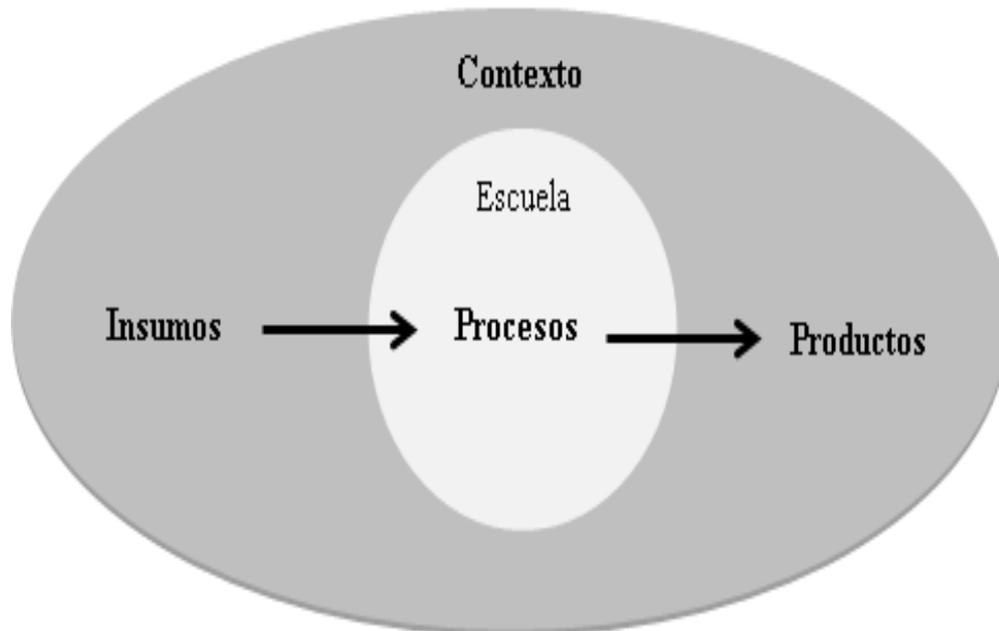
de los estudiantes y hacer seguimiento a la calidad educativa que se brinda en las instituciones educativas. Además, valorar los avances en un determinado lapso de tiempo, establecer el impacto de los programas y finalmente plantear acciones específicas de mejoramiento.

La Prueba Saber 11° evalúa las competencias de Lectura Crítica, Matemáticas, Ciencias Naturales, Sociales y Ciudadanas e Inglés¹². Su diseño está alineado con los Estándares Básicos de competencias establecidos por el Ministerio de Educación Nacional entendidos como referentes comunes a partir de los cuales es posible establecer que tanto los estudiantes y el sistema educativo en su conjunto está cumpliendo con unas expectativas de calidad en términos de lo que saben y saben hacer.

Sumado a esto, las pruebas permiten realizar estudios sobre las variables que más influyen en el rendimiento académico, y el ICFES con el fin de mejorar la calidad de la educación del país extendió el proceso de evaluación e inició el estudio de los factores asociados al rendimiento escolar, para este fin se aplicó un cuestionario de prueba en el año 2012 para recoger información sobre el contexto de los estudiantes, sus familias y la institución educativa. Para este estudio el ICFES implementó el modelo Contexto, Insumos, Procesos y Productos (CIPP) el cual permite seleccionar variables de un contexto particular para describir de forma clara y sencilla como éstas influyen en el rendimiento académico. Ver figura 1.

¹² Instituto Colombiano para la Evaluación de la Educación ICFES. (2016). El antes y ahora del examen

Figura 1. Modelo CIPP.



Fuente: Marco de factores asociados Saber 3°, 5° y 9°.

En el contexto se recogen todas las variables externas a la institución educativa las cuales se caracterizan por los aspectos: social, económico, político, cultural y familiar. Dentro de estos se incluyen la ubicación geográfica de la institución educativa, así como la modalidad, dependencia administrativa y tamaño del establecimiento. Otros aspectos relacionados con las características personales de los estudiantes y de sus hogares como el nivel socioeconómico, el género, la motivación y el auto concepto académico también hacen parte de esta categoría.

Los Insumos hacen referencia a los recursos con los que se cuenta en la institución educativa e incluye el historial académico de los estudiantes como un factor determinante del desempeño de los mismos. Dentro de los indicadores de infraestructura escolar están los servicios básicos, el acceso a computadoras y la conexión a internet, entre otras, el tiempo efectivo de aprendizaje y las estrategias de clasificación de estudiantes al interior de las sedes; las variables asociadas a los antecedentes escolares como la asistencia a educación preescolar y la repetición de grado también se incluyen en esta categoría.

En los procesos se tiene en cuenta las actividades y estrategias que se implementan en las instituciones educativas para que los estudiantes apropien los conocimientos y logren desarrollar las competencias que se establecen desde el MEN. Además, se encuentran aquellas variables que permiten

caracterizar el clima al interior del aula y de la institución educativa, la gestión de los directores y la satisfacción de los docentes y de todo el personal de la institución.

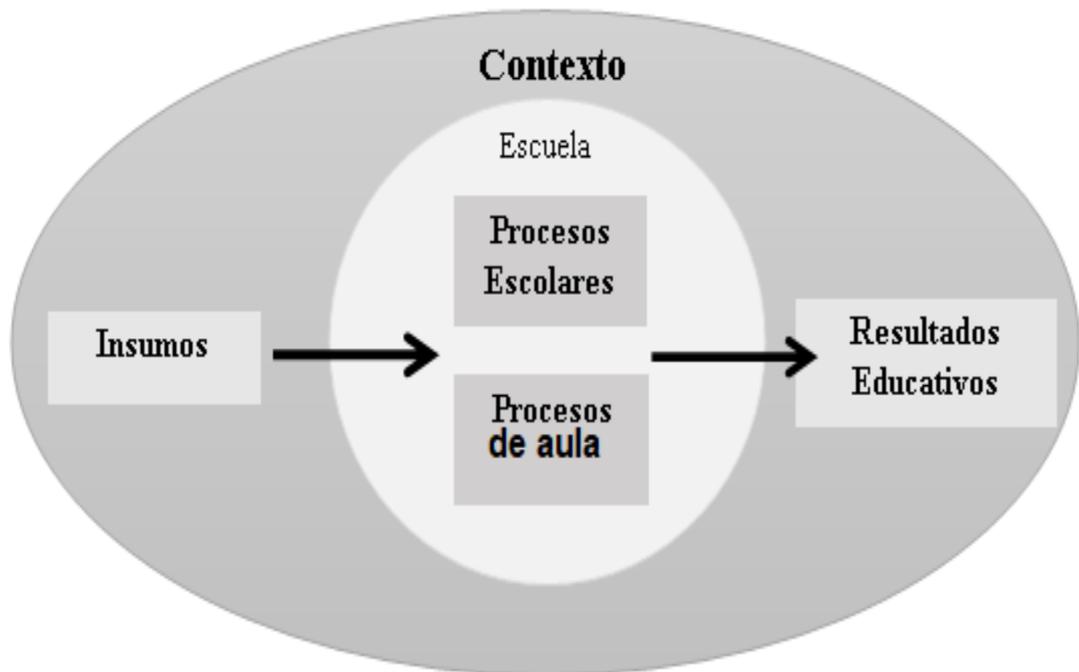
Finalmente, los productos son los resultados de las actividades empleadas en las instituciones y las políticas públicas, uno de esos resultados es el desarrollo cognitivo, social, emocional y ciudadano que conllevan a formar una persona integral.

El modelo CIPP cumple cuatro criterios generales que permiten hacer un juicio valorativo; Es completo puesto que incluye todas las categorías de los factores asociados al aprendizaje y la influencia de estas en el proceso educativo. Es claro, porque presenta resultados que se pueden entender a partir cualquier ámbito desde especialistas en investigación educativa, así como tomadores de decisiones, docentes y rectores. Presenta direccionalidad de las relaciones que se espera encontrar a partir del planteamiento de hipótesis sobre los factores asociados al aprendizaje. Por último, el modelo es flexible y permite incluir variables que son relevantes para el aprendizaje y desarrollo integral de los estudiantes, esto debido a que se pueden presentar cambios o transformaciones socioculturales.

Factores Asociados. Como se mencionó anteriormente el marco de factores asociados sigue el modelo CIPP, como se muestra en la figura 2, y dentro de cada una de las categorías incluidas en él se definen unas variables que han sido conceptualizadas y analizadas a nivel mundial y pueden usarse de forma dinámica para responder a políticas públicas o adaptarse a cualquier contexto a nivel nacional, regional y local. Para este estudio se tomó como referencia el Marco de factores asociados Saber 3°, 5° y 9° los cuales se definen a continuación¹³:

¹³ factores asociados

Figura 2. Marco de factores asociados.



Fuente: Marco de factores asociados Saber 3°, 5° y 9°.

Contexto. Los factores asociados incorporan las variables más cercanas al proceso educativo porque son estas las que caracterizan a los estudiantes e instituciones educativas, algunas de ellas son:

Características de los estudiantes. Según estudios de Coleman y la UNESCO, los aspectos sociodemográficos, económicos y culturales de los estudiantes y sus familias son variables importantes a la hora de explicar los resultados escolares. Estos factores a pesar de influir en el aprendizaje no se pueden modificar puesto que no dependen las instituciones educativas. Las variables que se consideran de mayor seguimiento dentro del marco de factores asociados son:

- **Género:** según la Unesco, Goldin y Niederle, se evidencia una brecha de desigualdad en el rendimiento académico de los niños y niñas puesto que se ven reflejados los estereotipos de género que se marcan desde la familia y la sociedad, por ejemplo, la gran desventaja que se muestra en las niñas frente al aprendizaje de las matemáticas y a los niños en lectura.

- **Edad:** esta variable se relaciona con el grado que cursa el estudiante; permite identificar aspectos como la repetición de año, la extra edad, la desescolarización parcial o deserción escolar. Según la ley general de educación un estudiante se encuentra en extra edad si supera en dos años la edad promedio del grado cursado; por tanto, la escolarización es obligatoria entre los 5 y 15 años de edad, desde primero a noveno grado.
- **Trabajo infantil:** es una variable que afecta de manera negativa al rendimiento escolar ya que las condiciones físicas de los niños y niñas disminuyen, pues no están preparados para asumir tareas laborales pesadas. Se debe diferenciar entre el trabajo infantil y la colaboración en tareas hogareñas puesto que el primero afecta negativamente el rendimiento escolar y el segundo muestra una relación positiva.
- **Nivel socioeconómico familiar:** según Célis, Jiménez & Jaramillo y el Icfes, la variable socioeconómica es explicativa frente al desempeño de los estudiantes en las Pruebas Saber, puesto que es una limitante para su desarrollo. Las condiciones del entorno donde viven, el nivel educativo de los padres, la situación laboral, la disposición de artefactos tecnológicos, entre otras, son algunos aspectos de interés ya que son un aliciente importante para el rendimiento escolar.
- **Violencia en el entorno del hogar:** un entorno violento tiene influencias negativas en cuanto al rendimiento escolar y la formación ciudadana. Para poder determinar el efecto que tiene la violencia en el entorno familiar y en el desempeño académico se debe hacer un estudio de la frecuencia con que ocurren dichas situaciones y en qué tipo de violencia se está incurriendo.

Involucramiento parental: el acompañamiento y participación de los padres en el proceso de formación académica beneficia el desarrollo cognitivo y no cognitivo de los hijos. Crea actitudes de motivación frente al aprendizaje, reflejadas en los resultados escolares, se fortalece el desarrollo motivacional y ayuda a crear hábitos de estudio en familia.

- **Pertenencia a un pueblo originario o etnia:** el desafío que hay en las instituciones educativas es el de incluir y atender a estudiantes que pertenecen a alguna etnia en particular; ya que estos presentan condiciones socioeconómicas bajas y de difícil acceso a la educación. Por esto, los establecimientos educativos deben apropiarse sus currículos para la flexibilización del aprendizaje y las necesidades que requieren por los estudiantes.

Necesidades educativas especiales: la política de una “Educación para todos” planteada por la UNESCO deja ver la importancia de tener en cuenta esta variable NEE dentro de los currículos de los establecimientos educativos del país, ya que incide significativamente en el desempeño académico de los estudiantes. Se deben distinguir entre necesidades educativas cognitivas y físicas, dado que en las primeras se debe buscar fortalecer y alcanzar los logros mínimos planteados para cada grado, mientras que las necesidades físicas se asocian más a la adecuación de espacios físicos que le permita la accesibilidad al establecimiento.

Auto-concepto académico: es entendido como las creencias que tienen los estudiantes sobre sus capacidades para hacer bien sus labores académicas y obtener buenos resultados. La confianza que tenga cada estudiante sobre su desempeño afecta de manera positiva o negativa sus resultados; esto se evidencia en el cumplimiento de tareas, la participación en clases y la asistencia a las mismas. En la formación del auto-concepto académico se tiene en cuenta aspectos como la comparación personal, la comparación social, la percepción de los padres, profesores y compañeros.

Motivación: esta variable muestra una estrecha relación con el desempeño académico de los estudiantes es así como Ryan and Deci proponen dos tipos de motivación: la intrínseca cuando se refiere a la motivación causada por gusto e interés propio y la extrínseca cuando es causada por algún incentivo en particular. En el marco conceptual de los cuestionarios PISA, la motivación es reconocida como parte de los comportamientos o actitudes positivas frente al aprendizaje y cuando se logra controlar la variable socioeconómica la motivación intrínseca tiene resultados positivos frente al aprendizaje. Además, la motivación en el ámbito educativo se debe entender como la manera en que las relaciones entre los estudiantes y sus profesores entran en juego para facilitarla o limitarla con el fin de lograr un aprendizaje significativo, razón por la cual el profesor debe favorecer la denominada motivación intrínseca, que lleva a la curiosidad y el descubrimiento de lo nuevo, a diferencia de la motivación extrínseca en la que el estudiante es movido por otros para realizar diferentes actividades.

- **Estrategias de aprendizaje:** las estrategias de aprendizaje son las que facilitan la adquisición del conocimiento y la adquisición de nuevas habilidades. Existen diferentes estrategias de aprendizaje, pero se intensifican los estudios en las tradicionales y las cooperativas. Caracterizándose las primeras por el uso de la memorización como recurso de aprendizaje y las segundas por la creación colectiva del conocimiento.

Características de las escuelas: en este apartado se analizan aquellas variables que hacen referencia a las instituciones educativas. Dichas características están referenciadas en los cuestionarios que se aplican a los estudiantes al momento de presentar las Pruebas Saber. Entre las variables más relevantes en esta sección están:

Zona en la que se ubica la institución educativa (urbana o rural): sobre esta variable se han realizado estudios que evidencian que las instituciones que se encuentran más aisladas poseen menos posibilidades de contar con docentes idóneos, materiales y recursos tecnológicos, apoyo técnico y se da una mayor deserción estudiantil, principalmente por la falta de recursos económicos de los padres que en muchas ocasiones emplean a sus hijos para realizar las labores del campo. Además de esto, en diversos establecimientos que se encuentran en las zonas rurales existen modos de organización diferentes pues es común encontrar escuelas; multigrado o unidocente.

En este sentido la ubicación de la institución educativa es un factor que cobra importancia para explicar el desempeño académico de los niños y niñas del país y se puede medir diferenciando entre escuelas rurales y urbanas, el municipio, el departamento, la zona y la entidad territorial en la que se encuentra.

- **Dependencia administrativa de la escuela:** la influencia ejercida por el tipo de dependencia es una de las variables que cobra mayor importancia dentro de las características de las escuelas, ya que algunos autores como Chubb consideran que los entes privados exigen por parte de los directivos y docentes el mejoramiento continuo en los aprendizajes de los estudiantes por la presión ejercida por los padres. Sin embargo, en la práctica se ha encontrado que dicha afirmación en pocos casos se cumple, indicando que la educación privada por sí sola no asegura mejores aprendizajes y oportunidades para los estudiantes. Es importante dentro de este aspecto distinguir dos dependencias administrativas: públicas y privadas, las primeras son controladas por el estado, las segundas están regidas por algún ente privado, además se pueden diferenciar por la orientación religiosa, la procedencia de los aportes o los llamados colegios en concesión donde se entrega la administración de un colegio que cuenta con infraestructura pública a uno privado.
- **Tamaño de la institución educativa y el aula de clases:** frente a este factor existen contraposiciones de diversos autores frente al tamaño de la institución y el número de estudiantes en el aula; Crenshaw y Lamdin afirman que la cantidad de estudiantes atendidos en la institución educativa influye en el tipo de enseñanza y en el proceso de aprendizaje de los estudiantes, sin embargo Ajani & Akinyele y Stevenson se enfocan más en el tamaño relativo de la institución afirmando que el número de docentes por estudiante explica de mejor manera los efectos en el apoyo del proceso de enseñanza.

Otros factores que se tienen en cuenta para explicar este factor hacen referencia a la capacidad de la planta docente, por la cantidad de cursos y

niveles, y por el tipo de educación: técnica profesional, artística, científica, entre otros.

- **Nivel socioeconómico promedio por aula y establecimiento:** este factor se puede determinar a partir del estrato de la institución y toma importancia al identificar el efecto de los pares en cada estudiante, el cual es beneficioso en el desarrollo escolar pues se da la heterogeneidad social y académica de los estudiantes, a nivel de aula y escuela, además impacta el desarrollo del currículum, la calidad de las interacciones y la capacidad del profesor para desarrollar procesos de enseñanza efectivos. Algunas características asociadas al nivel socioeconómico tienen que ver con el ingreso, posesión de bienes, nivel de estudios de la madre y/o padre, ocupación de la madre y/o padre.
- **Violencia en el entorno de la institución educativa:** según Bowen & Bowen y la Unesco (2015b) esta característica evidencia las situaciones de conflicto y por ende de procesos de marginación al interior de las instituciones educativas. Las repercusiones al respecto recaen en principalmente en los estudiantes, frustrando su adecuado desarrollo académico y posteriormente los bajos desempeños en Pruebas Saber.

Insumos. Los insumos o recursos educativos con los que cuenta la institución educativa son una parte importante en los resultados académicos; ya que a través de estos los estudiantes adquieren un aprendizaje significativo que fortalece sus competencias escolares. A continuación, se presentan aspectos que se relacionan con esta categoría:

- **Antecedentes escolares:** es importante reconocer que los estudiantes traen aprendizajes y conocimientos previos que influyen de manera directa en el rendimiento académico. Factores asociados como la participación en preescolar y la repetición de grado son antecedentes que traen consigo los estudiantes y se consideran favorables para la adquisición de nuevas competencias en la formación académica.
- **Repetición de grado:** este tema ha sido estudiado por diferentes autores como Holmes y Roderick quienes indican que la repetición de año tiene efectos negativos a nivel académico y emocional a través del tiempo. Cuanto menor sea el grado de repetición mayores serán los efectos o consecuencias negativas para el aprendizaje. En Latinoamérica, la repitencia de grado en la institución educativa ha sido identificada como la variable que más influencia en los logros de aprendizaje, luego del nivel socioeconómico de los estudiantes, lo cual refleja un impacto negativo como medida efectiva

para mejorar el desempeño académico. En la actualidad Colombia es uno de los países de Latinoamérica con las tasas de repitencia más bajas en educación primaria; sin embargo, al considerar las consecuencias de esto, resulta conveniente revisar con mayor detalle las políticas del sistema educativo colombiano.

- **Asistencia a educación preescolar:** según la OECD la educación preescolar tiene efectos positivos sobre los resultados escolares y en otros ámbitos como la reducción de la pobreza y la movilidad social; por tal razón la asistencia al preescolar es favorable para el desarrollo de la sociedad en general. En este sentido, los gobiernos deben enfocarse en desarrollar programas que fortalezcan y apoyen este grado buscando equidad, calidad y cumplimiento con los estándares mínimos que les permitan a los niños y niñas gozar de todos los beneficios que se puedan ofrecer en pro de la calidad en el aprendizaje. Además de esto, la vinculación a programas de educación preescolar ofrece características diferenciadoras en las trayectorias escolares de los niños y niñas; quienes asisten a educación preescolar formal llegan mejor preparados para ajustarse a las exigencias escolares de la educación básica y media ofrecida en los distintos establecimientos del país.
- **Uso de tecnologías:** hoy en día la sociedad esta movida por las Tecnologías de la Información y la Comunicación (TIC's) pues brindan múltiples posibilidades de acceder a información de manera más rápida y efectiva, razón por la cual incluir dentro de la práctica docente el uso de las TIC's, reducidas al uso de computadoras, dispositivos electrónicos y el acceso a internet, es de vital importancia dentro de los procesos de enseñanza y aprendizaje. Uno de los aportes de las TIC's es el aprendizaje colaborativo que puede darse; a nivel interinstitucional entre docentes y estudiantes y a nivel intrainstitucional facilitando la colaboración y el aprendizaje colectivo entre familias y centros educativos alrededor del mundo, y permitir el acceso a procesos formativos virtuales para todos los ciudadanos.

Según los resultados obtenidos en las pruebas PISA 2012 se evidencia pequeños efectos en los resultados, situación que se explica por la poca formación de docentes en cuanto a las capacidades de entendimiento y pensamiento frente a estrategias con el uso de las tecnologías y de estudiantes y sus inexistentes prácticas de pedagogía encaminadas hacia uso intensivo y adecuado de las TIC.

Infraestructura y equipamiento escolar: la relación que existe entre la infraestructura y el rendimiento académico presenta una relación positiva ya

que; disponer de una biblioteca, de computadores o tecnologías de información y comunicación, de espacios recreativos, de salones de clase adecuados para todos los estudiantes y de servicios sanitarios adecuados, le da la posibilidad a los estudiantes de contar con ambientes propicios que generan mayor interés y motivación por el aprendizaje.

Calificación docente y conocimiento profesional: un buen docente se caracteriza por tres aspectos que lo hacen idóneo en su desempeño laboral: los criterios de selección inicial, la calidad de instrucción y la capacitación continua. Además, es importante que el docente conozca las características de sus estudiantes y busque estrategias pedagógicas que le certifiquen el aprendizaje de sus alumnos independientemente de sus características de origen.

Materiales educativos: hace referencia a los recursos o herramientas que permiten la profundización de los conceptos y la apropiación del conocimiento tanto de estudiantes como docentes.

Algunos elementos que hacen parte de los materiales educativos son los artefactos tecnológicos, los libros de texto, la disponibilidad de internet, los útiles escolares, entre otros. El uso y aprovechamiento que se les dé a estos recursos en el aula de clases durante el desarrollo de las sesiones pedagógicas define en gran medida los resultados esperados frente al logro educativo.

- **Programa de útiles escolares gratuitos:** si bien es cierto este tipo de programas son de vital importancia para apoyar los procesos de aprendizaje de los estudiantes, ya que les permite tener acceso y disponibilidad de herramientas educativas para la adquisición de conocimientos, también se sabe que estos programas tienen criterios de focalización (colegios o estudiantes más desventajados socioeconómicamente) sin embargo los materiales que entregan a los establecimientos educativos muchas veces no son necesariamente los que se requieren para orientar el aprendizaje, pues como se mencionó anteriormente no se cuenta con la instrucción adecuada para que el uso adecuado de dichas herramientas.
- **Tiempo efectivo dedicado al aprendizaje:** este componente hace referencia al tiempo que se dedica exclusivamente al aprendizaje haciendo diferencia entre las horas de estudio de los estudiantes y el tipo de jornada al que asisten. Otros factores que permiten identificar el tiempo efectivo dedicado al aprendizaje son la inasistencia escolar y la puntualidad por parte de estudiantes y docentes, la frecuencia con la que ocurren interrupciones durante la clase, debido a un ambiente de indisciplina nocivo para el aprendizaje o a factores de distracción externos al control del docente como

ruido o condiciones climáticas inapropiadas, y el manejo y organización del tiempo por parte del docente.

Procesos. Los procesos que ocurren en la institución educativa y en los salones de clase representan el núcleo de la labor educativa, puesto que a través de las interacciones cotidianas entre docentes y estudiantes se promueve el aprendizaje (ICFES, 2016). Los constructos que hacen parte de los procesos educativos se definen a continuación:

- **Liderazgo educativo:** esta es una de las variables que más influye en el aprendizaje de los estudiantes dado el papel que juega el desarrollo de programas, proyectos y prácticas educativas. Los roles de liderazgo educativo orientan principalmente al logro del aprendizaje y se focalizan en los aspectos de la gestión instruccional. El liderazgo instruccional es un factor importante para orientar las metas y focalizar los agentes internos hacia los resultados académicos en los procesos de enseñanza. Algunos de los procesos escolares transformadores que deben ser tomados en cuenta por las personas a cargo del liderazgo educativo son: empoderamiento de los profesores, fortalecimiento de la visión institucional y de objetivos compartidos, y capacidades de rendición de cuentas.

Desarrollo y colaboración profesional para mejorar la enseñanza: el trabajo cooperativo y colaborativo entre maestros permite mejorar el aprendizaje y el desempeño escolar; el intercambio de experiencias sobre las prácticas de aula, conlleva a mejorar las propias y contribuye a desarrollar habilidades en los estudiantes.

Además de esto, el trabajo colaborativo permite fortalecer la planificación, el desarrollo, la evaluación y la retroalimentación para mejorar las prácticas educativas.

- **Clima escolar:** hace alusión al conjunto de normas y acuerdos que se establecen en la institución educativa para propiciar un ambiente acorde a las necesidades de los estudiantes y garantizar un aprendizaje efectivo. Un clima laboral y escolar positivo se asocia con los logros académicos. Algunos constructos que se relacionan con esta variable son:
- **Presencia efectiva de normas y acuerdos:** un ambiente de aprendizaje armonioso, basado en el respeto con normas claras y el cumplimiento de las mismas, potencia el aprendizaje estudiantil; contrario a esto, cuando se presentan ambientes hostiles, de autoritarismo e irrespeto por parte del docente o los estudiantes, se genera indisciplina y pérdida de tiempo valioso que podría dedicarse a la adquisición de conocimientos.

- **Relaciones interpersonales:** entre los aspectos a tener en cuenta en las relaciones interpersonales están la comunicación efectiva y asertiva entre los actores educativos, así como las relaciones basadas en la confianza. Es impotente hacer énfasis en la actitud y disposición que tienen los docentes para resolver las dudas de sus estudiantes, la actitud de los estudiantes frente a los procesos de enseñanza y el tipo de comunicación que se da en clase.
- **Percepción sobre el aula y la institución educativa:** en la literatura se argumenta que existen creencias frente a las experiencias vividas por los alumnos en la escuela que afecta significativamente su desempeño académico y su comportamiento. En este sentido la práctica docente cobra la mayor importancia y debe estar basada en aspectos como; claridad en los objetivos planteados para el desarrollo de las clases de manera que los estudiantes comprendan el sentido de las actividades planteadas, la escucha de inquietudes y el apoyo frente a adversas situaciones que se les puedan presentar, preguntar sobre cómo se sienten en la institución y propiciar relaciones que permitan favorecer la permanencia y el éxito escolar.
- **Prácticas de enseñanza docente:** este aspecto hace énfasis en la calidad instruccional del profesor en donde se evidencia la interacción entre este y sus estudiantes, generando una relación positiva si se cuenta con un conjunto de habilidades por parte del docente, entre las cuales están en primer lugar el desarrollo de conceptos, la estimulación y el desarrollo de habilidades de pensamiento superior, la integración de conceptos y de conocimientos previos, la producción propia y las relaciones y aplicaciones al mundo real. En segundo lugar, la calidad de la retroalimentación dada por parte del docente a los estudiantes, donde el profesor debe desarrollar bases claras para que el estudiante logre avanzar en su aprendizaje, que haga seguimiento constante, que promueva la meta cognición y que reafirme positivamente a los estudiantes. Y, en tercer lugar, la modelación lingüística, que se refiere a la capacidad del profesor para extender y desarrollar constantemente el lenguaje en los estudiantes, a través de conversaciones frecuentes, preguntas abiertas y el uso de lenguaje avanzado, entre otros.

Resultados Educativos. Actualmente la expectativa de los sistemas escolares, frente al objetivo de promover el desarrollo integral de los estudiantes, se ha convertido en la base fundamental para propender por un buen desempeño académico; partiendo del aprendizaje en las disciplinas académicas, el desarrollo de competencias ciudadanas y el bienestar subjetivo de los estudiantes. Por esta razón es importante describir cada una de ellas como se hace a continuación.

- **Aprendizaje en disciplinas académicas:** según lo establecido en los estándares básicos de competencias es de vital importancia tener un criterio claro y público que permita juzgar si un estudiante, una institución o el sistema educativo en su conjunto cumplen con unas expectativas comunes de calidad, a esto se define estándar y partiendo de esta definición cobra importancia la medición de los aprendizajes en las distintas disciplinas académicas de las cuales no solo hacen parte las áreas básicas como matemáticas y lenguaje, sino también aquellas que forman parte del desarrollo integral del estudiante y su formación como ciudadanos.
- **Ciudadanía:** la formación ciudadana hace referencia, de forma general, a aprender a vivir con otros y a participar activamente de una sociedad democrática. En este sentido, este factor asociado se debe tener en cuenta como primordial en la formación académica y ciudadana de los estudiantes, siendo las instituciones educativas el lugar propicio para la socialización y el desarrollo de habilidades y actitudes en relación a la vida en democracia, el respeto por los otros, el cumplimiento de las normas acordadas, participar y expresar su opinión, entre otros aspectos.
- **Bienestar subjetivo:** a pesar de ser un factor muy importante en la mediciones de resultados académicos muchas veces quedan relegadas a un segundo plano, debido a que las variable de tipo psicológico asociadas a aspectos socioemocionales no presentan características generales que permitan una medición; el bienestar subjetivo se puede entender como la percepción de bienestar individual, que hace parte de la inteligencia emocional como la capacidad que tiene una persona para reconocer los sentimientos propios y ajenos y por lo tanto es inteligente o hábil para el manejo de los sentimientos. En este sentido se debe tener en cuenta este factor dentro del proceso de aprendizaje de los estudiantes.

Estructura y alineación. A partir del segundo semestre del año 2014 las pruebas fueron alineadas para establecer comparativos con los diferentes exámenes que se aplican como son las Pruebas Saber 3°, 5°, 9°, 11° y Saber Pro, con el fin de monitorear el progreso que han tenido los estudiantes después de aplicar la prueba en un determinado grado, por ejemplo, un estudiante de primaria que presentó la prueba Saber 3° después de dos años aplica la prueba Saber 5° con los resultados obtenidos de las dos pruebas se puede establecer un comparativo y hacer un análisis del proceso de formación que alcanzó en la Básica Primaria¹⁴.

Las Pruebas Saber miden las competencias, es decir, pretenden indagar a los estudiantes cómo piensan y cómo utilizan el saber adquirido en los diferentes contextos de aprendizaje. Se entiende por competencia al conjunto de habilidades, destrezas y conocimientos que desarrolla una persona para comprender, transformar y participar de manera activa y crítica en el mundo que lo rodea.

Actualmente, el Ministerio de Educación Nacional (MEN) concibe el objetivo de la educación como el desarrollo de determinadas competencias y en consecuencia a estas como el objeto de la evaluación. Dentro de las diferentes competencias que pueden desarrollarse a lo largo del proceso educativo hay una categoría que merece especial atención: la de las competencias genéricas, entendidas como aquellas que resultan indispensables para el desempeño social, laboral y cívico de todo ciudadano, independientemente de su oficio o profesión. Contrastan con las competencias (no genéricas) propias de oficios o actividades laborales particulares, que resultan de un entrenamiento especializado.

De lo anterior, las competencias genéricas son la base más importante en la formación que reciben los estudiantes de la educación básica; puesto que no solamente son indispensables para subsistir en la vida cotidiana, sino que también le permite al individuo tener una visión exitosa en cuanto a la participación en el mercado laboral, en las relaciones interpersonales en su contexto, con su familia, en procesos democráticos que aporten un bien para la construcción de la paz.

Aunado a lo anterior, la Prueba Saber 11° evalúa las competencias de Matemáticas, Lectura Crítica, Sociales y Ciudadanas, Ciencias Naturales e Inglés; descritas a continuación.

¹⁴ Alineación del examen SABER 11°

Lectura Crítica. Esta prueba evalúa 3 competencias, las cuales son necesarias para comprender, interpretar y evaluar textos. Las tres competencias evaluadas se describen a continuación.

Identificar y entender los contenidos locales que conforman un texto: esta competencia consiste en la capacidad de comprender el significado de las palabras, expresiones y frases que aparecen explícitamente en el texto. La evaluación de esta competencia no indaga por conocimientos propios de la gramática, pero sí por la comprensión de las relaciones semánticas que existen entre los diferentes elementos que constituyen una frase. En ausencia de esta competencia, no es posible contar con las dos siguientes.

Comprender cómo se articulan las partes de un texto para darle un sentido global: esta competencia consiste en la capacidad de comprender cómo se relacionan formal y semánticamente los elementos locales que constituyen un texto, de manera que este adquiera un sentido global (a nivel del párrafo, sección, capítulo, etcétera). En esa medida, las preguntas correspondientes a esta competencia involucran varios elementos locales de un texto (por ejemplo, diferentes frases o párrafos) y exigen reconocer y comprender su articulación. Esta competencia es necesaria para contar con la siguiente.

Reflexionar a partir de un texto y evaluar su contenido: esta competencia consiste en la capacidad de enfrentar un texto críticamente. Incluye evaluar la validez de argumentos, advertir supuestos, derivar implicaciones, reconocer estrategias argumentativas y retóricas, relacionar contenidos con variables contextuales, etcétera. Se trata de la competencia propiamente crítica y, como tal, exige un ejercicio adecuado de las dos anteriores. La prueba hace énfasis en la segunda y en la tercera competencia puesto que, de acuerdo con los estándares básicos de competencias en lenguaje del MEN (2006), al finalizar undécimo grado, los estudiantes deben poder leer de manera crítica

Matemáticas. Esta prueba evalúa 3 competencias las cuales son necesarias para enfrentar distintas situaciones o contextos que pueden ser resueltos, justificados, transformados; Con el uso de algunas herramientas matemáticas. Las tres competencias evaluadas se describen a continuación.

Interpretación y representación: esta competencia consiste en la habilidad para comprender y transformar la información presentada en formatos distintos como tablas, gráficas, conjuntos de datos, diagramas, esquemas, etcétera, así como la capacidad de utilizar estas representaciones para extraer información relevante que permita, entre otras cosas, establecer relaciones matemáticas e identificar tendencias y patrones. Con el desarrollo de esta competencia se espera que un estudiante utilice coherentemente registros como el simbólico, el natural, el gráfico y todos aquellos que se dan en situaciones que involucran las matemáticas. Esta competencia se relaciona con el proceso de comunicación, representación y razonamiento, definidos en los estándares básicos de competencias.

Formulación y ejecución: esta competencia se relaciona con la capacidad de plantear y diseñar estrategias que permitan solucionar problemas provenientes de diversos contextos, bien sean netamente matemáticos o bien sean aquellos que pueden surgir en la vida cotidiana, siempre que sean susceptibles de un tratamiento matemático. Se relaciona también con la habilidad o destreza para seleccionar y verificar la pertinencia de soluciones propuestas a determinados problemas y estrategias de solución desde diferentes puntos de vista. Con el desarrollo de esta competencia se espera que un estudiante diseñe estrategias apoyadas en herramientas matemáticas, proponga y determine rutas posibles para la solución de problemas, siga estrategias dadas para encontrar soluciones y, finalmente resuelva las situaciones que se le propongan. Esta competencia evalúa el proceso de formulación, tratamiento y resolución de problemas; el proceso de formulación, comparación y ejercitación de procedimientos, y el proceso de modelación, todos descritos en los estándares básicos de competencias.

Argumentación: esta competencia se relaciona con la capacidad para validar o refutar conclusiones, estrategias, soluciones, interpretaciones y representaciones en diversas situaciones, siempre justificando el por qué o el cómo se llegó a estas, a través de ejemplos y contraejemplos, o señalando y reflexionando sobre inconsistencias presentes. Con el desarrollo de esta competencia se espera que un estudiante justifique la aceptación o el rechazo de afirmaciones, interpretaciones y estrategias de solución basado en propiedades, resultados o verbalizando procedimientos matemáticos. Cabe indicar que esta competencia se relaciona con los procesos de razonamiento y la modelación definidos en los estándares básicos de competencias.

Sociales y competencias ciudadanas. Evalúa la habilidad para analizar distintos eventos, conceptos, posturas, modelos, dimensiones y contextos, así como la capacidad de reflexionar y emitir juiciosos críticos sobre estos. También los conocimientos y habilidades que permiten al estudiante comprender el mundo social.

Pensamiento social: esta competencia evalúa la capacidad del estudiante para usar conceptos básicos de las ciencias sociales que permiten la comprensión de problemáticas y fenómenos sociales, políticos, económicos, culturales y geográficos, como también los principios básicos de la Constitución, y la estructura y el funcionamiento del sistema político colombiano. De esta manera, por un lado, con las preguntas de esta competencia el estudiante se enfrenta a diferentes situaciones sociales que deberá analizar a través de conceptos básicos de las ciencias sociales. Hacen parte de esta competencia la comprensión de dimensiones temporales y espaciales de eventos y prácticas sociales; la habilidad del estudiante para analizar las dimensiones históricas de eventos o problemáticas, y la capacidad de relacionar sucesos o procesos con contextos geográficos y sus características, es decir, de conectar eventos e identificar razones que justifiquen esas conexiones. Por otro lado, con esta competencia se evalúa el

conocimiento que el estudiante tiene sobre los fundamentos del modelo de Estado social de derecho y sus particularidades en nuestro país, los derechos y deberes ciudadanos establecidos en la Constitución, la organización del Estado, las funciones y los alcances de las diferentes ramas del poder y de los organismos de control y el marco básico que fundamenta la participación ciudadana en Colombia.

Interpretación y análisis de perspectivas: esta competencia hace referencia a la capacidad del estudiante para pensar analítica y críticamente sobre las problemáticas y fenómenos sociales. A través de esta competencia se evalúan principalmente dos habilidades: el reconocimiento de diversas opiniones, posturas e intereses y el análisis crítico de fuentes y argumentos.

La primera habilidad hace referencia a la capacidad del estudiante de analizar una problemática según las perspectivas de diferentes actores sociales o colectivos involucrados en ella. Las preguntas correspondientes a esta habilidad requieren que los estudiantes, a propósito de un conflicto, estén en capacidad de comprender su origen, entender qué buscan los diferentes actores, identificar las coincidencias y diferencias entre sus intereses, y valorar la reacción (de aceptación o rechazo) de las partes ante una propuesta de solución.

La segunda habilidad da cuenta de la capacidad del estudiante de evaluar fuentes primarias y secundarias, corroborar y contextualizar la información que suministran, y valorar la pertinencia y solidez de los enunciados o argumentos que exponen. Se espera que el estudiante valore la credibilidad de las fuentes y analice los argumentos y perspectivas que allí se presentan. Si bien existen diferentes formas para evaluar el uso o la información proveniente de una fuente, una de ellas consiste en valorar la experticia de un autor para hablar de un tema. También se busca que el estudiante pueda inscribir una fuente en un contexto económico, político o cultural y pueda considerar otras fuentes para determinar si respaldan o contradicen la información que la fuente inicial presenta.

Respecto a esta segunda habilidad, además se espera que los estudiantes que culminan grado 11° estén en capacidad de evaluar qué tan fuertes son las razones aducidas para apoyar una conclusión presentada en una fuente, si las evidencias que apoyan una conclusión son hechos objetivos (por ejemplo, investigaciones versus opiniones personales), o si un argumento tiene sesgos, prejuicios o generalizaciones no justificadas.

Esta competencia resulta fundamental para el ejercicio de la ciudadanía en la medida en que demuestra la habilidad del estudiante para ponderar argumentos que conduzcan, por ejemplo, a decisiones tales como seleccionar

un candidato a partir de su programa de gobierno, o el análisis de problemáticas sociales o económicas a través de la lectura crítica del editorial de un periódico de gran relevancia nacional.

Pensamiento reflexivo y sistémico: esta competencia se entiende como la capacidad de comprender la realidad social desde una perspectiva sistémica. Es decir, de reconocer distintas formas de aproximarse a los problemas, de identificar las complejas relaciones que los configuran, de adelantar procesos de indagación flexibles y reflexivos, y de adoptar posturas críticas frente a los usos sociales de las ciencias sociales. Para ello, el estudiante debe contar con dos habilidades: identificar modelos conceptuales que orientan decisiones sociales y establecer relaciones entre las distintas dimensiones presentes en las problemáticas sociales y en sus posibles alternativas de solución.

La primera habilidad hace referencia al conocimiento de modelos conceptuales y de sus posibles usos en determinados contextos. Así, hacen parte de esta la identificación de las características propias de un modelo, de los principios que lo sustentan y de valorar decisiones como coherentes o incoherentes con los objetivos o postulados del modelo.

La segunda habilidad involucra la capacidad de los estudiantes para identificar distintas dimensiones o ámbitos de la vida social, y para analizar los posibles efectos que tendría una medida o intervención sobre tales dimensiones. Esto implica el reconocimiento de las relaciones entre los ámbitos (social, económico, ambiental y político, entre otros) que están presentes en un problema o decisión. Se considera que para entender una situación problemática es necesario identificar sus dimensiones y reconocer que en cada una puede haber diferentes valores o ideas que se privilegian en un momento dado. Se espera entonces que el estudiante comprenda que cada situación involucra diferentes dimensiones y que, en ocasiones, pueden contraponerse entre sí. Por ende, debe comprender que las decisiones que se toman pueden favorecer unas dimensiones e ir en detrimento de otras.

Ahora bien, cada una de las competencias evaluadas en la prueba de Sociales y ciudadanas cuenta con dos afirmaciones. Las afirmaciones son “enunciados que se realizan acerca de los conocimientos, habilidades y capacidades que se pretende inferir a partir de las respuestas dadas por los estudiantes en las pruebas” es decir, se trata de descripciones de aquello que se espera que un estudiante esté en capacidad de hacer cuando ha desarrollado una competencia.

Ciencias naturales. Esta prueba evalúa la capacidad para construir significativamente el conocimiento existente, aprendiendo a razonar, a resolver

problemas, a tomar decisiones, a pensar con rigurosidad y a valorar de manera crítica el conocimiento y su efecto en la sociedad y en ambiente.

Uso comprensivo del conocimiento científico: es la capacidad de comprender y usar nociones, conceptos y teorías de las ciencias naturales en la solución de problemas, y de establecer relaciones entre conceptos y conocimientos adquiridos, y fenómenos que se observan con frecuencia.

Explicación de fenómenos: es la capacidad de construir explicaciones y comprender argumentos y modelos que den razón de fenómenos, y de establecer la validez o coherencia de una afirmación o de un argumento relacionado con un fenómeno o problema científico.

Indagación: vincular a los estudiantes con la forma como se amplía y modifica el conocimiento científico es esencial para formar ciudadanos alfabetizados científicamente. Esta competencia, que en la estructura de la prueba abarca un 40 % del total de preguntas, se define como la capacidad para comprender que, a partir de la investigación científica, se construyen explicaciones sobre el mundo natural. Además, involucra los procedimientos o metodologías que se aplican para generar más preguntas o intentar dar respuestas a estas. El proceso de indagación en ciencias incluye, entre otras cosas, observar detenidamente la situación planteada, formular preguntas, recurrir a libros u otras fuentes de información, hacer predicciones, plantear experimentos, identificar variables, realizar mediciones y organizar y analizar resultados. En el aula de clases no se trata de que el alumno repita un protocolo ya establecido o elaborado por el docente, sino que el estudiante formule sus propias preguntas y diseñe su propio procedimiento.

Inglés. Esta prueba evalúa las habilidades comunicativas a nivel de lectura y uso de lenguaje en esta lengua en relación con el Marco Común Europeo de Referencia para las lenguas (MCER).

Esta prueba está conformada por 7 partes las cuales cada una evalúa una habilidad específica de la lengua inglesa.

Parte 1: En esta parte de la prueba se indaga por el conocimiento pragmático del estudiante. En particular, el estudiante debe reconocer el propósito comunicativo de un aviso y el lugar donde este puede aparecer, según el propósito. Para ello, el estudiante debe decidir en qué sitio puede encontrar los avisos que aparecen inicialmente. En cada pregunta hay 3 opciones de respuesta, A, B o C, de las cuales deberá marcar solo una en su hoja de respuestas.

Parte 2: Esta parte evalúa el conocimiento lexical del estudiante. Para ello, el estudiante debe comprender una serie de descripciones con el fin de relacionarlas con una lista de palabras. Por tanto, el estudiante debe buscar la relación entre una lista de ocho palabras clasificadas de la letra A, a la letra H y

las preguntas de esta parte, las cuales describen una de las ocho palabras de la lista. Cabe indicar que hay más palabras (A a H) de las que el estudiante necesita.

Parte 3: La parte 3 evalúa al estudiante en su conocimiento comunicativo. En concreto, el estudiante debe elegir la intervención más adecuada que un interlocutor 1 haría frente a lo dicho por un interlocutor 2. Lo anterior se relaciona con la pertinencia del lenguaje usado en situaciones particulares recreadas en pequeñas conversaciones. En consonancia con lo anterior, en esta parte el estudiante debe completar conversaciones cortas, seleccionando la respuesta correcta de las tres opciones de la hoja respuesta.

Parte 4: Con base en un texto, en esta parte se evalúa el conocimiento gramatical del estudiante. Puntualmente, se trata de elegir las palabras más adecuadas para completar un texto. Para ello, el estudiante debe leer y prestar atención a una serie de espacios, puesto que, para cada uno de ellos, debe seleccionar la palabra correcta entre las tres opciones, A, B o C, en su hoja de respuestas.

Parte 5: En esta parte de la prueba, el estudiante debe realizar un ejercicio de comprensión de lectura literal de un texto. Este ejercicio consiste en seleccionar la paráfrasis que permite responder correctamente a cada pregunta planteada sobre un texto. El estudiante debe seleccionar la respuesta correcta, para cada pregunta, entre tres opciones, A, B o C, en su hoja de respuestas.

Parte 6: Con base en un texto, el estudiante debe llevar a cabo un proceso de lectura inferencial. En esta parte se plantean distintas preguntas sobre la intención del autor y los aspectos generales y particulares destacables del texto. El estudiante debe seleccionar la respuesta correcta, para cada pregunta, entre cuatro opciones, A, B, C o D, en su hoja de respuestas.

Parte 7: a partir del texto que se presenta, en esta parte de la prueba se evalúa el conocimiento tanto gramatical como lexical del estudiante. El ejercicio consiste en elegir las palabras más adecuadas para completar el texto. Para ello, el estudiante debe seleccionar la palabra correcta entre las cuatro opciones, A, B, C o D, en su hoja de respuestas, que completa cada uno de los espacios del texto presentado.

Los resultados obtenidos ofrecen información sobre los desempeños de los estudiantes que presentaron dicha evaluación en un conjunto de áreas. La Prueba Saber 11° permite la producción de diferentes reportes de resultados los cuales son:

- Reportes individuales.
- Reportes por establecimiento educativo, con los resultados de cada estudiante.
- Clasificación de planteles según categorías de rendimiento.

- Reportes de resultados históricos por establecimiento educativo.
- Reportes de resultados históricos por municipio.
- Reportes de resultados históricos por departamento.
- Informes nacionales de resultados.

Además, de entregar este tipo de reportes de resultados, el ICFES presenta informes de resultados en los cuales se analizan las diferencias en los desempeños alcanzados con respecto a variables tales como nivel socioeconómico, género, tipo de Institución Educativa (oficial o privada), zona de ubicación (urbana o rural), municipios, departamentos o regiones, entre otras. Estos análisis tienen como propósito dar elementos para la toma de decisiones en cuanto a políticas y programas de mejoramiento de la calidad de la educación. El ICFES pone a disposición de los investigadores las bases de datos con esta información, para que esta se utilice en el desarrollo de nuevos estudios¹⁵.

Bodega de datos

De acuerdo a¹⁶ una bodega de datos (Data Warehouse - DW) es un repositorio de datos central; el cual, abarca todo el contexto del problema. Para Inmon¹⁷, es una colección de datos orientados a un determinado ámbito (empresa, organización, etc.), integrado, no volátil y variable en el tiempo, que soporta la toma de decisiones en la organización en la que se utiliza.

Su expansión se ha dado en la medida en que las organizaciones han migrado de esquemas orientados solo a la recolección de datos, a esquemas enfocados en el análisis de los mismos. Son implementados en la Inteligencia de Negocios por su capacidad de ejecutar varias consultas analíticas.

Las principales características que tiene la bodega de datos son¹⁸:

- La información debe ser fácilmente accesible.
- La información debe ser consistente.
- La bodega de datos debe ser adaptada a múltiples cambios.
- Los datos en ellas son generalmente de carácter histórico.
- Deben ser seguras para proteger la información.

¹⁵ tipos de resultados generados a partir de las evaluaciones

¹⁶ Big Data Fundamentals. Concepts, Drivers & Techniques

¹⁷ Relational and Dimensional Techniques, Wiley Publishing, 2003.

¹⁸ The data warehouse toolkit

Mercado de datos

Los mercados de datos están conformados por un subconjunto de los datos en la bodega¹⁹. Comúnmente, estos mercados de datos pertenecen a un dominio específico del problema, en términos empresariales, a un departamento o un área específica de la empresa; la diferencia con la bodega de datos es su alcance, esta cubre toda la organización (ver figura 3).

Existen dos tipos de mercados de datos²⁰:

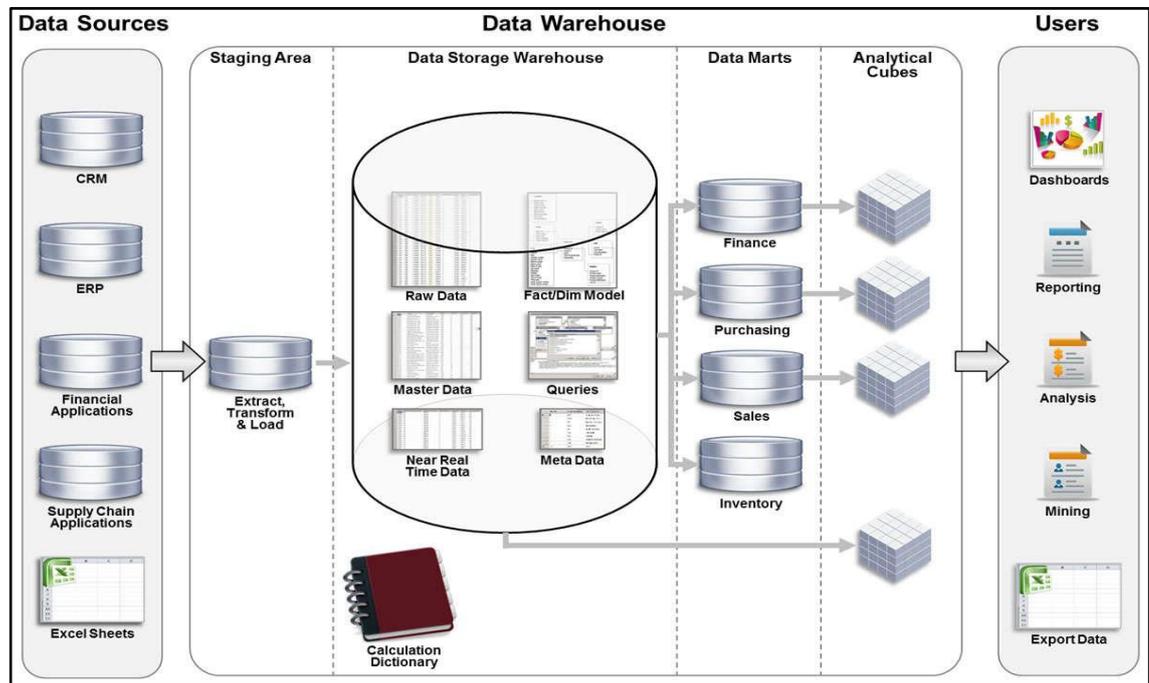
Mercados de datos dependientes: Es aquel que toma sus datos a partir de la bodega de datos directamente; por lo tanto, su información es consistente y ya está integrada.

Mercados de datos independientes: Es aquel que toma sus datos de fuentes externas a la bodega de datos, es por ello que sus datos pueden no estar integrados ni ser consistentes. Típicamente los datos son extraídos por medio de bases de datos tipo (OLTP) o con almacenes operacionales de datos (ODS).

¹⁹ Big Data Fundamentals. Concepts, Drivers & Techniques

²⁰ Data Mart Consolidation

Figura 3. Estructura lógica de una bodega de datos.



Fuente: The Data Warehouse Group (<http://www.data-warehouse.com.au>).

Diseño de un mercado de datos

Existen múltiples alternativas para el diseño de un mercado de datos, entre los cuales se encuentran los aportes de Kimball y Rose²¹, el trabajo de Malinowski y Zimányi²² o la propuesta de Inmon²³.

La metodología de Ralph Kimball presenta un enfoque Bottom-up (de mayor a menor o ascendente), lo cual se traduce la implementación de mercados de datos en áreas específicas de la organización, y poco a poco irlos integrando en una bodega de datos Corporativa.

Rivadeneira en²⁴ describe los cuatro principios básicos que fundamentan el ciclo de vida de la metodología Kimball:

²¹ The data warehouse toolkit

²² Advanced data warehouse design

²³ Building the Data Warehouse

²⁴ La metodología de Kimball para el diseño de almacenes de datos

Centralizarse en el negocio: Es fundamental identificar los requerimientos del negocio y su valor asociado.

Construir una infraestructura de información adecuada: La herramienta diseñada debe ser fácil de usar, de alto rendimiento, integrada y que refleje los requerimientos del negocio.

Realizar entregas en incrementos significativos: Construir la bodega de datos en incrementos entregables; es una concepción similar a las metodologías ágiles de construcción de software.

Ofrecer la solución completa: Construir los elementos necesarios para entregar valores a los usuarios: una bodega de datos sólida (probado y accesible), herramientas de análisis y de consulta, generación de reportes, capacitación, soporte y documentación.

Esta metodología de desarrollo se compone de fases, como se muestra en la Figura 4, en la cual es importante destacar dos aspectos:

El rol central de la definición de requerimientos, que construyen el soporte inicial de las actividades subsiguientes además de su influencia en la construcción del plan de proyecto.

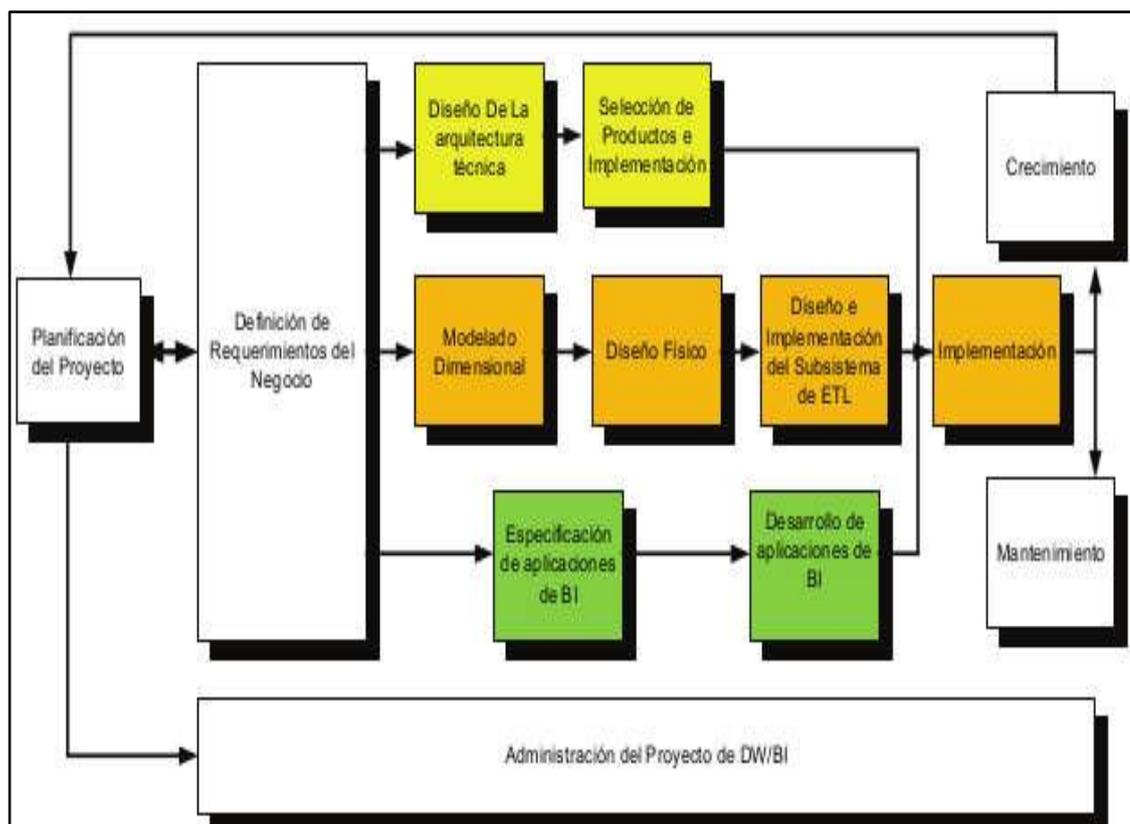
Se identifican tres rutas o caminos enfocados en diferentes áreas:

Tecnología (Camino Superior): Involucra análisis de software y hardware.

Datos (Camino del Medio): Diseño e implementación de bodega de datos.

Aplicación de Inteligencia de Negocios (Camino Inferior): Diseño e implantación de aplicaciones para los usuarios finales.

Figura 4. Fases de la metodología Kimball.



Fuente: Ralph Kimball, The Data Warehouse Lifecycle Toolkit.

Análisis multidimensional

Hace referencia a las diferentes perspectivas para analizar los datos almacenados en una bodega de datos; las tablas de dimensiones contienen los atributos que se quieren estudiar según el nivel de detalle. En un modelo dimensional se pueden encontrar los siguientes componentes²⁵:

Dimensiones: Representan las áreas temáticas, las líneas o los sujetos del negocio; brindan una línea base para estructurar la información mediante el uso de un grupo de atributos.

Atributos: Agrupación de elementos o ítems relacionados dentro de una dimensión que tiene el mismo nivel lógico.

Jerarquías: Son representadas por un ordenamiento lógico dentro de la dimensión, se encuentran formadas por los diferentes tipos de relaciones entre los atributos de una misma dimensión.

²⁵ Desarrollo de un datamart de información académica de estudiantes de la Escuela de Ciencias y Sistemas de la Facultad de Ingeniería de la USAC

Tablas de Hechos: Un hecho es un indicador de negocios que da sentido al análisis de las dimensiones. Las tablas de hechos involucran indicadores relacionados con un proceso específico del negocio y las llaves o claves de las dimensiones que intervienen en dicho proceso.

PostgreSQL

PostgreSQL es un sistema de gestión de bases de datos objeto-relacional, distribuido bajo una licencia similar a la BSD (Berkeley Software Distribution) y con su código fuente disponible libremente. Cuenta con más de 15 años de desarrollo activo y una arquitectura probada que se ha ganado una sólida reputación por su fiabilidad, integridad de datos y corrección. Se ejecuta en todos los sistemas operativos; incluyendo Linux, UNIX, Mac OS X, Solaris y Windows.

PostgreSQL utiliza un modelo cliente/servidor y es compatible con el almacenamiento de grandes objetos binarios, incluyendo imágenes, sonidos, o de vídeo. El desarrollo de PostgreSQL no es manejado por una empresa y/o persona, sino que es dirigido por una comunidad de desarrolladores que trabajan de forma desinteresada²⁶.

PENTAHO

La plataforma Open Source Pentaho Business Intelligence cubre muy amplias necesidades de análisis de los datos y de los informes empresariales. Las soluciones de Pentaho están escritas en java. Eso hace que Pentaho sea una solución muy flexible para cubrir una amplia gama de necesidades empresariales, tanto como las típicas como las sofisticadas y específicas al negocio²⁷.

Los módulos de la plataforma Business Intelligence son:

Reporting: Un módulo de los informes ofrece la solución adecuada a las necesidades de los usuarios. Pentaho Reporting es una solución basada en el proyecto JFreeReport el cual permite generar informes ágiles y de gran capacidad. Pentaho Reporting permite la distribución de los resultados del

²⁶ PostgreSQL

²⁷ Qué es Pentaho y para qué sirve

análisis en múltiples formatos. Todos los informes incluyen la opción de imprimir o exportar a formato PDF, XLS, HTML, y texto. Los reportes Pentaho permiten también programación de tareas y ejecución automática de informes con una determinada periodicidad.

Análisis: PENTAHO Análisis suministra a los usuarios un sistema avanzado de análisis de información. Con uso de las tablas dinámicas (pivot tables, crosstabs), generadas por Mondrian y JPivot, el usuario puede navegar por los datos, ajustando la visión de los datos, los filtros de visualización, añadiendo o quitando los datos de agregación. Los datos pueden ser representados en una forma de SVG o flash, o también integrados con los sistemas de minería de datos y los portales web (portlets). Además, con Microsoft Excel Analysis Services, se puede analizar los datos dinámicos en Microsoft Excel (usando la conexión a OLAP server Mondrian).

Dashboards: Todos los componentes del módulo, Pentaho Reporting y Pentaho Análisis, pueden formar parte de un dashboard. En Pentaho Dashboards es muy fácil incorporar una gran variedad en tipos de gráficos, tablas y velocímetros (dashboard widgets) e integrarlos con los Portlets JSP, en donde podrá visualizar informes, gráficos y análisis OLAP.

Data Mining: El análisis en Pentaho se realiza con una herramienta Weka. La integración de datos se realiza con una herramienta Kettle ETL (Pentaho Data Integration) que permite implementar los procesos ETL.

Modelo multidimensional

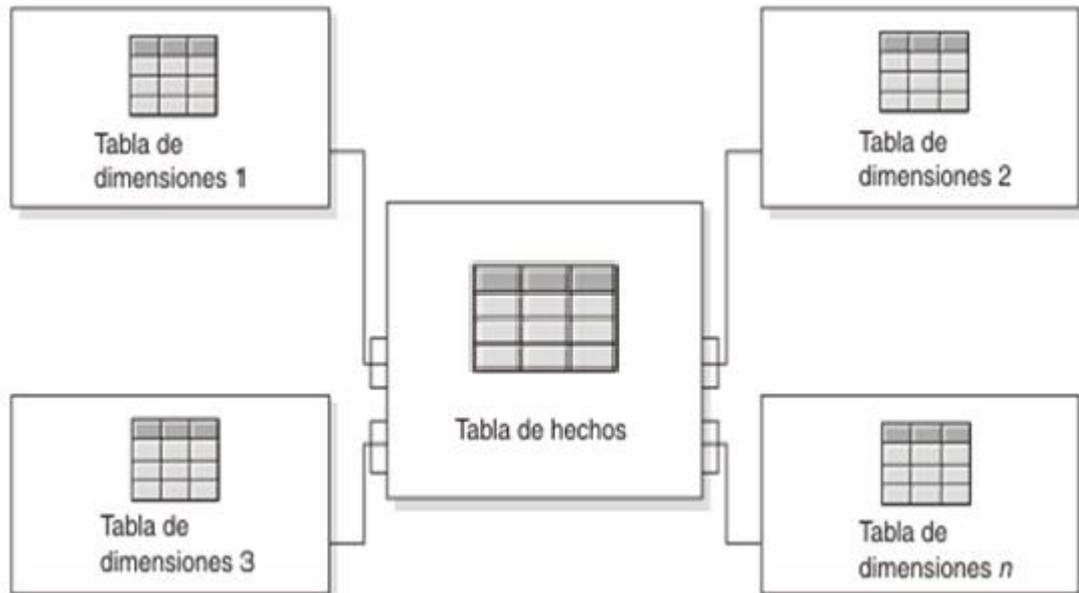
Existen principalmente dos tipos de esquemas para estructurar los datos como un almacén de datos.

- **Modelo multidimensional en estrella**

Un esquema de estrella es una base de datos relacional que consta de una sola tabla de hechos central rodeada de tablas de dimensiones.

En la figura 5, se muestra un esquema de estrella con una sola tabla de hechos y cuatro tablas de dimensiones. Un esquema de estrella puede tener cualquier número de tablas de dimensiones. Las ramas situadas al final de los enlaces que conectan las tablas indican una relación de muchos a uno entre la tabla de hechos y cada tabla de dimensiones.

Figura 5. Modelo multidimensional en estrella.

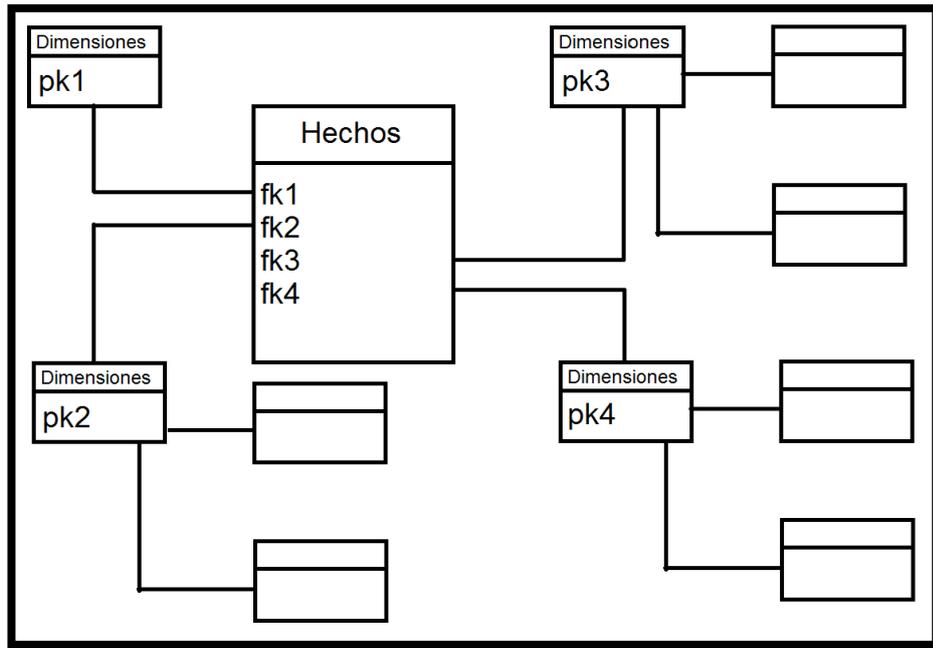


Fuente: Esta investigación

- **Modelo multidimensional copo de nieve**

En la figura 6, modelo multidimensional copo de nieve se deriva del esquema en estrella, en el que las tablas de dimensiones se normalizan en múltiples tablas. Por esta razón, la tabla de hechos deja de ser la única tabla del esquema que se relaciona con otras tablas, y aparecen nuevas uniones.

Figura 6. Modelo multidimensional copo de nieve



Fuente: Esta investigación

Operadores OLAP

Drill Y Roll: Las operaciones de agregación (DRILL) y disgregación (ROLL) se pueden hacer sobre:

- Atributos de una dimensión sobre los que sea definido una jerarquía: DRILL-DOWN, ROLL-UP.
- Sobre dimensiones independientes: DRILL-ACROSS, ROLL-ACROSS.

Pivot: Esta operación reorienta las dimensiones en el informe.

Slice (Proyectar):

- Reduce la dimensionalidad del esquema eliminando alguna dimensión.
- Aplicar esta operación implica pérdida de detalle en los hechos (se aumenta la granularidad) por lo que se tendrá que utilizar operadores de agregación para la obtención de los nuevos.
- Definición algebraica:

$$\pi(C_S^n, D_1, \dots, D_m, D_{m+1} = d_{m+1}, \dots, D_n = d_n) = C_S^m \quad m \leq n$$

Dice (Restricción):

- Restringe los valores en las dimensiones según algunas condiciones.
- No se modifica la estructura del cubo en cuanto a las dimensiones y niveles de estas.

- No se modifica el nivel de detalle de los hechos, pero si su número, dado que aquellos que tuvieran como coordenadas valores no considerados deben ser eliminados.
- Definición algebraica: $\sigma(C_s^m, P) = C_s^m$

Subregión de Obando

- **Ubicación:** Está ubicada al Sur de Nariño y la integran los municipios de Ipiales, Aldana, Guachucal, Cumbal, Cuaspud Carlosama, Pupiales, Puerres, Córdoba, Potosí, El Contadero, Iles, Gualmatán, Funes; Posee una extensión de 4.894 kilómetros cuadrados aproximadamente, que equivalen al 14.07% del área total del Departamento.
- **Población:** Está conformada por 273.776 habitantes que corresponden al 16.49% del total del Departamento; de los cuales 124.646 están ubicados en el sector urbano y 149.130 en el sector rural (50% son hombres y 50% mujeres). Etnográficamente está compuesto por 92.059 indígenas y 268 afrocolombianos.
- **Actividades económicas:** En el sector agropecuario se destaca el cultivo de papa, maíz, trigo, cebada, frijol, alverja; igualmente es significativa la explotación de ganado bovino, porcino, caballar y especies menores. También cabe resaltar la actividad comercial y artesanal²⁸.

²⁸ Ex-Provincia de Obando

1 IMPLEMENTACION

Para el desarrollo del proyecto se siguieron las fases propuestas por Kimball, en lo que él denominó “Ciclo de Vida Dimensional Del Negocio” (Business Dimensional Lifecycle). Las tareas realizadas en cada una de las fases son descritas a continuación.

1.1 Análisis de requerimientos

Como se mencionó anteriormente, ni en la Subregión de Obando, ni en otro municipio perteneciente al departamento Nariño se han desarrollado proyectos de investigación que almacenen y analicen los resultados de las Pruebas Saber 11° que presentaron los estudiantes de las Instituciones Educativas de la región; Los estudios que se han realizado en la Subregión con respecto al análisis de resultados de dichas pruebas, han sido de manera aislada en cada Institución, sin tener en cuenta la información que estás generarían al hacer un tratamiento más complejo de todos los datos. Lo cual es posible con la construcción de un mercado de datos.

El interés de este proyecto se centró en la construcción de un mercado de datos que almacene los datos históricos de los estudiantes que presentaron la Prueba Saber 11° de la Subregión de Obando, de tal forma que mediante una herramienta de Inteligencia de Negocios permita desarrollar análisis multidimensional y así generar distintos tipos de beneficios:

- Disponer de un repositorio de los resultados históricos de las Pruebas Saber 11° del período comprendido entre los años 2012 al 2016.
- Establecer indicadores de evaluación de la calidad.
- Mejorar la gestión institucional.
- Caracterizar a los estudiantes.
- Analizar el desempeño histórico de los estudiantes en las Pruebas Saber 11°.
- Apoyarse en un soporte real para generar políticas y estrategias para fortalecer los programas académicos.

1.2. Modelo dimensional

Para la construcción del mercado de datos se cuenta inicialmente con datos de 3'356.299 estudiantes de diferentes Instituciones académicas de diversas zonas del país y un total de 174 atributos, que en términos generales pueden ser agrupados en las siguientes categorías.

- Características generales del estudiante.
- Características socio-económicas del núcleo familiar del estudiante.
- Características de la infraestructura del lugar de residencia del estudiante.
- Características de la institución de educación media.
- Desempeño y resultados del estudiante en las diferentes pruebas que conforman la Prueba Saber 11°.

Partiendo de lo anterior, se establece las dimensiones y la granularidad para el mercado de datos que se presenta en la Tabla 1:

Tabla 1. Dimensiones y granularidad del mercado de datos.

| Dimensión | Granularidad |
|------------------------------|--|
| Estudiantes | Estudiante |
| Instituciones | Institución de Educación Media |
| Lugares | Municipio |
| Prueba Ciencias Naturales | Prueba ciencias naturales del estudiante |
| Prueba Inglés | Prueba ingles del estudiante |
| Prueba Lectura Crítica | Prueba lectura crítica del estudiante |
| Prueba Matemáticas | Prueba matemáticas del estudiante |
| Prueba Sociales y Ciudadanas | Prueba sociales y ciudadanas de estudiante |
| Tiempo | Semestre |

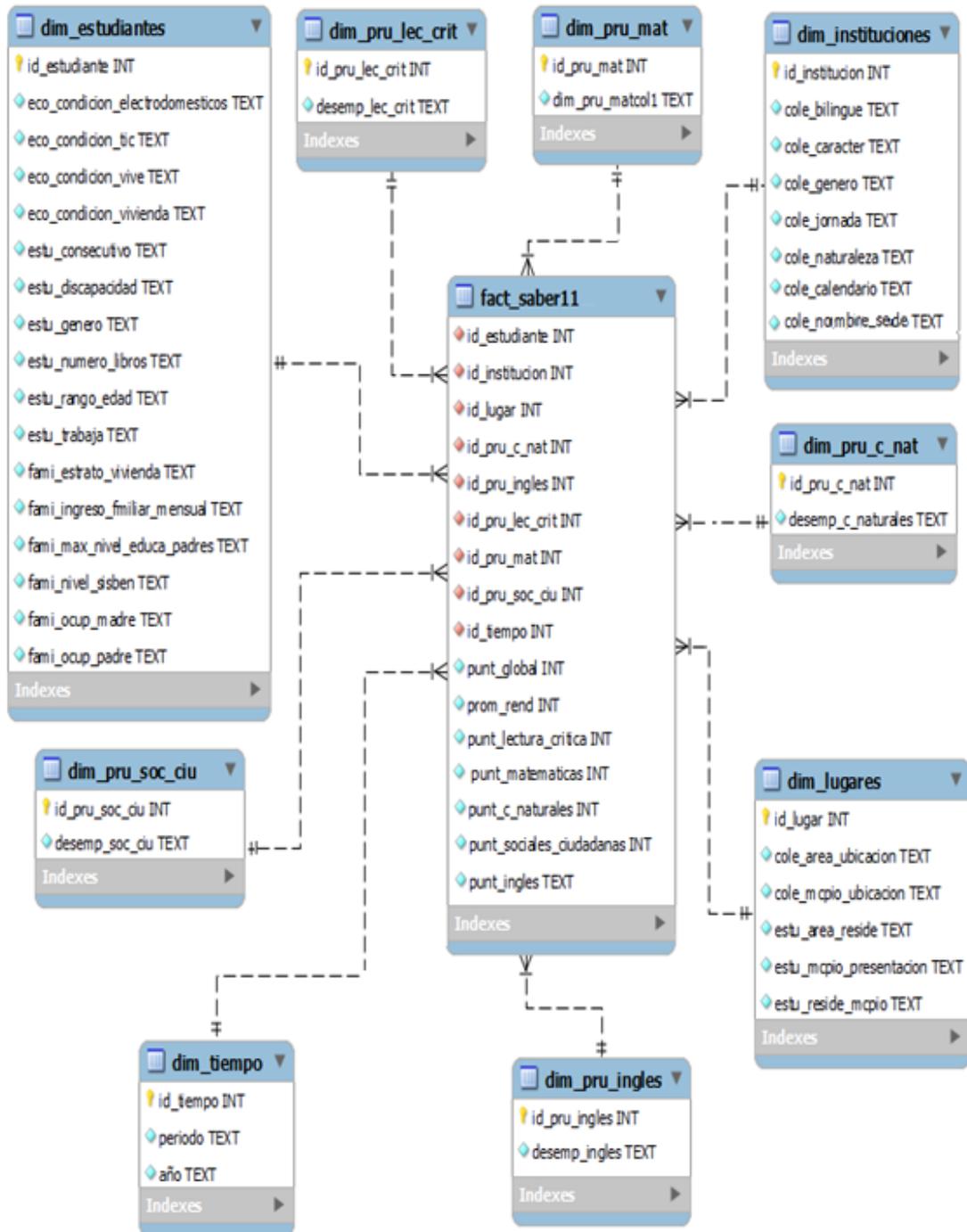
Se tiene una tabla de hechos (fact_saber11) que tiene como gránulo el estudiante, y se identifican las siguientes medidas sobre los hechos:

- ¿Cuántos estudiantes presentaron las Pruebas Saber 11° por municipio?
- ¿Cuántos estudiantes presentaron las Prueba Saber 11° por año?
- ¿Cuántos estudiantes presentaron la Prueba Saber 11° por rango de edades?

- ¿Cuántos estudiantes presentaron la Prueba de sociales y ciudadanas por rango de edades y desempeño?
- ¿Cuáles son los promedios de desempeño de la Prueba de sociales ciudadanas, ciencias naturales y matemáticas discriminados por rango de edad?
- ¿Cuáles son los promedios de desempeño de la Prueba de sociales ciudadanas, ciencias naturales y matemáticas describiendo por municipios?
- ¿Cuántos estudiantes presentaron la Prueba Saber 11° por condición TIC?
- ¿Cuántos estudiantes presentaron la Prueba de sociales y ciudadanas discriminado por condición TIC y desempeño de la prueba?
- ¿Cuáles son los promedios de la Prueba de sociales ciudadanas, matemáticas y ciencias naturales discriminado por condición TIC y promedio?
- ¿Cuántos estudiantes presentaron la prueba por género?
- ¿Cuántos estudiantes presentaron la prueba por año y género?
- ¿Promedio de la Prueba de sociales y ciudadanas en el período del 2012 al 2016 discriminado por género?
- ¿Cuántos estudiantes presentaron la Prueba de sociales y ciudadanas en el período del 2012 al 2016 discriminado por desempeño y género?
- ¿Cuántos estudiantes presentaron la Prueba de sociales y ciudadanas por desempeño?
- ¿Cuántos estudiantes presentaron la Prueba de sociales y ciudadanas por desempeño?
- ¿Cuántos estudiantes presentaron la Prueba Saber 11° en el período del 2012 al 2016 discriminado por estrato de vivienda?

Para la implementación del modelo multidimensional asociado al mercado de datos se decidió usar un esquema de tipo estrella el cual se presenta en la Figura 7.

Figura 6. Modelo multidimensional de tipo estrella.



Fuente: Esta investigación

1.2.1. Diccionario de datos. En las Tablas comprendidas de 2 a 11 se detalla el diccionario de datos asociados a cada una de las tablas del modelo multidimensional.

Tabla 2. Diccionario de datos de la dimensión estudiante.

| Nombre del Atributo | Tipo de Dato | Descripción |
|---------------------------------|---------------------|---|
| id_estudiante | Serial | Llave primaria |
| eco_condicion_electrodomesticos | Text | Condiciones de electrodomésticos en el hogar del estudiante |
| eco_condicion_tic | Text | Condiciones del uso de TIC en el hogar del estudiante |
| eco_condicion_vive | Text | Condiciones de vida del estudiante |
| eco_condicion_vivienda | Text | Condiciones de la vivienda del estudiante |
| estu_consecutivo | Text | Identificador del estudiante |
| estu_discapacidad | Text | Discapacidad o no |
| estu_genero | Text | Género |
| estu_numero_libros | Text | Cantidad de libros |
| estu_rango_edad | Text | Rango de edad del estudiante en el momento de presentar la prueba |
| estu_trabaja | Text | Si estudiante trabaja o no |
| fami_estrato_vivienda | Text | Estrato socioeconómico del estudiante |
| fami_ingreso_fmiliar_mensual | Text | Ingresos mensuales familiares en salarios mínimos |
| fami_max_nivel_educa_padres | Text | Máximo nivel educativo completo entre el padre y la madre |
| fami_nivel_sisben | Text | Nivel del Sisbén |
| fami_ocup_madre | Text | Ocupación de la madre |
| fami_ocup_padre | Text | Ocupación del padre |

Tabla 3. Diccionario de datos de la dimensión instituciones.

| Nombre del Atributo | Tipo de Dato | Descripción |
|----------------------------|---------------------|---|
| id_institucion | Serial | Llave primaria |
| cole_bilingue | Text | La institución educativa es bilingüe |
| cole_caracter | Text | Carácter académico de la institución |
| cole_genero | Text | Población del colegio |
| cole_jornada | Text | Jornada de la institución |
| cole_naturaleza | Text | Naturaleza del colegio oficial y no oficial |
| cole_nombre_sede | Text | Nombre de la institución |
| cole_calendario | Text | Calendario institución |

Tabla 4. Diccionario de datos de la dimensión lugares.

| Nombre del Atributo | Tipo de Dato | Descripción |
|-------------------------|--------------|--|
| id_lugar | Serial | Llave primaria |
| cole_area_ubicacion | Text | Ubicación de la institución; Urbana-Rural |
| cole_mcpio_ubicacion | Text | Ubicación de la institución por municipios |
| estu_area_reside | Text | Lugar donde reside el estudiante; Urbana-Rural |
| estu_mcpio_presentacion | Text | Municipio donde el estudiante presento la prueba |
| estu_reside_mcpio | Text | Municipio donde el estudiante reside |

Tabla 5. Diccionario de datos de la dimensión tiempo.

| Nombre del Atributo | Tipo de Dato | Descripción |
|---------------------|--------------|---|
| id_tiempo | Serial | Llave primaria |
| Periodo | Text | Año y semestre en el que presentó la prueba |
| Año | Text | Año en el que presento la prueba |

Tabla 6. Diccionario de datos de la dimensión prueba ciencias naturales.

| Nombre del Atributo | Tipo de Dato | Descripción |
|---------------------|--------------|--|
| id_pru_c_nat | Serial | Llave primaria |
| desemp_c_naturales | Text | Desempeño, descripción cualitativa de las habilidades y conocimientos que podrían tener si se ubican en determinado nivel. |

Tabla 7. Diccionario de datos de la dimensión prueba inglés.

| Nombre del Atributo | Tipo de Dato | Descripción |
|---------------------|--------------|--|
| id_pru_ingles | Serial | llave primaria |
| desemp_ingles | Text | Desempeño, descripción cualitativa de las habilidades y conocimientos que podrían tener si se ubican en determinado nivel. |

Tabla 8. Diccionario de datos de la dimensión prueba lectura crítica.

| Nombre del Atributo | Tipo de Dato | Descripción |
|---------------------|--------------|---------------------------------------|
| id_pru Lec_crit | Serial | Llave primaria |
| desemp Lec_crit | Text | Desempeño, descripción cualitativa de |

| | | |
|--|--|--|
| | | las habilidades y conocimientos que podrían tener si se ubican en determinado nivel. |
|--|--|--|

Tabla 9. Diccionario de datos de la dimensión prueba matemáticas.

| Nombre del Atributo | Tipo de Dato | Descripción |
|---------------------|--------------|--|
| id_pru_mat | Serial | Llave primaria |
| desemp_lec_crit | Text | Desempeño, descripción cualitativa de las habilidades y conocimientos que podrían tener si se ubican en determinado nivel. |

Tabla 10. Diccionario de datos de la dimensión prueba sociales ciudadanas.

| Nombre del Atributo | Tipo de Dato | Descripción |
|---------------------|--------------|--|
| id_pru_soc_ciu | Serial | Llave primaria |
| desemp_soc_ciu | Text | Desempeño, descripción cualitativa de las habilidades y conocimientos que podrían tener si se ubican en determinado nivel. |

Tabla 11. Diccionario de datos de la tabla de hechos.

| Nombre del Atributo | Tipo de Dato | Descripción |
|---------------------|--------------|-------------------------------|
| id_estudiante | Integer | Llave foránea |
| id_institucion | Integer | Llave foránea |
| id_lugar | Integer | Llave foránea |
| id_pru_c_nat | Integer | Llave foránea |
| id_pru_ingles | Integer | Llave foránea |
| id_pru_lec_crit | Integer | Llave foránea |
| id_pru_mat | Integer | Llave foránea |
| id_pru_soc_ciu | Integer | Llave foránea |
| id_tiempo | Integer | Llave foránea |
| punt_global | Integer | Puntaje global |
| prom_rend | Integer | Promedio puntaje global |
| punt_c_nat | Integer | Puntaje ciencias naturales |
| punt_ingles | Integer | Puntaje ingles |
| punt_lec_crit | Integer | Puntaje lectura critica |
| punt_mat | Integer | Puntaje matemáticas |
| punt_soc_ciu | Integer | Puntaje sociales y ciudadanas |

1.3. Diseño del sistema de extracción, transformación y carga (ETL)

Para implementar el mercado de datos se utilizó el motor de base de datos relacional PostgreSQL versión 1.22.2 donde se utilizaron 10 repositorios descargados de la página oficial del ICFES, cada uno con la información referente a las pruebas saber 11° de los años 2012, 2013, 2014, 2015 y 2016.

A continuación, se describe el proceso de Extracción, Transformación y Carga para construir y poblar el mercado de datos.

1.3.1 **Fase de extracción.** En esta fase se construye un repositorio de datos limpio y estandarizado a partir de los 10 repositorios fuentes, para tal fin se hizo una serie de procesos descritos a continuación:

- **Integración de datos:** Se creó el repositorio saber11_2012_2016 utilizando toda la información de los 10 repositorios fuentes descargados de la página oficial del ICFES. Este repositorio cuenta inicialmente con datos de 3'356.299 estudiantes de diferentes Instituciones académicas de diversas zonas del país y un total de 174 atributos.

En la Tabla 12 se presenta el número de registros que contiene cada repositorio desde el 2012 al 2016.

Tabla 12. Número de registros de cada repositorio.

| Repositorio | No. De registros |
|------------------------|------------------|
| saber11_2012_1 | 95946 |
| saber11_2012_2 | 575646 |
| saber11_2013_1 | 86469 |
| saber11_2013_2 | 573661 |
| saber11_2014_1 | 88684 |
| saber11_2014_2 | 574394 |
| saber11_2015_1 | 108257 |
| saber11_2015_2 | 573127 |
| saber11_2016_1 | 74126 |
| saber11_2016_2 | 605989 |
| Total registros | 3.356.299 |

Posteriormente a la creación de este repositorio base, se hace un minucioso análisis de los atributos, con la ayuda de los diccionarios de datos que facilita el ICFES para su mejor interpretación, se crea un nuevo diccionario de datos para el repositorio creado. A continuación, en la tabla 13 se presenta el diccionario de datos con su respectiva descripción de cada atributo.

Tabla 13. Diccionario de datos.

| cod | VARIABLES | DESCRIPCION DE VARIABLES |
|-----|----------------------------|--|
| 1 | COLE_AREA_UBICACION | Ubicación del colegio tiene dos valores urbana y rural |
| 2 | COLE_BILINGUE | La institución educativa es bilingüe |
| 3 | COLE_CALEDARIO | Calendario del colegio |
| 4 | COLE_CARACTER | Carácter del colegio académico comercial |
| 5 | COLE_COD_DANE_INSTITUCION | Código asignado a la institución por el Dane |
| 6 | COLE_COD_ICFES | Código ICFES asignado a la institución |
| 7 | COLE_COD_MCPIO_UBICACION | Código Dane del municipio donde está ubicado el colegio |
| 8 | COLE_DEPTO_UBICACION | Departamento donde está ubicado el colegio |
| 9 | COLE_GENERO | Población del colegio descripción por géneros |
| 10 | COLE_JORNADA | Jornada del colegio |
| 11 | COLE_MCPIO_UBICACION | Ubicación del plantel en el municipio |
| 12 | COLE_NATURALEZA | Naturaleza del colegio oficial y no oficial |
| 13 | COLE_NOMBRE_SEDE | Nombre de la sede en la que estudia el estudiante |
| 14 | COLE_VALOR_PENSION | Valor de la pensión pagada por el estudiante en el último año |
| 15 | DECIL_C_NATURALES | Decil prueba de ciencias naturales |
| 16 | DECIL_COMP_CIUDADANA | Decil prueba de competencias ciudadanas |
| 17 | DECIL_INGLES | Decil prueba de prueba de ingles |
| 18 | DECIL_LLECTURA_CRITICA | Decil prueba de lectura crítica |
| 19 | DECIL_MATEMATICAS | Decil prueba de matemáticas |
| 20 | DECIL_RAZONA_CUANT | Decil prueba de razonamiento cuantitativo |
| 21 | DECIL_SOCIALES_CIUDADANAS | Decil prueba de sociales y ciudadanas |
| 22 | DECIL_C_NATURALES | Decil prueba de ciencias naturales |
| 23 | DESEMP_COMP_FLEXIBLE | Desempeño en componente flexible |
| 24 | DESEMP_INGLES | Desempeño en ingles según las bandas del marco |
| 25 | DESEMP_LLECTURA_CRITICA | Desempeño en lectura crítica |
| 26 | DESEMP_MATEMATICAS | Desempeño matemáticas |
| 27 | DESEMP_SOCIALES_CIUDADANAS | Desempeño sociales y ciudadanas |
| 28 | ESTU_ANO_MATRICULA_PRIMERO | En qué año se matriculó por primera vez en el 1° grado de primaria |
| 29 | ESTU_ANO_MATRICULA_SEXTO | En qué año terminó y aprobó el 6° grado de primaria |
| 30 | ESTU_ANO_TERMINO_BACHILL | Año en el que el estudiante termino el bachillerato. Solo aplica para INDIVIDUALES |
| 31 | ESTU_ANO_TERMINO_QUINTO | En qué año terminó y aprobó el 5° grado de primaria |
| 32 | ESTU_ANOS_COLEGIO_ACTUAL | Cuántos alumnos hay en total en el curso (salón) |
| 33 | ESTU_ANOS_PREESCOLAR | Cuántos años de educación preescolar cursó |
| 34 | ESTU_AREA_RESIDE | Área donde vive el inscrito |
| 35 | ESTU_CARRDESEADA_TIPO | Aspiración de carrera profesional a estudiar |

Tabla 13. (Continuación)

| | | |
|----|-------------------------------|---|
| 36 | ESTU_COD_MCPIO_PRESENTACION | Código del municipio de aplicación del examen según DIVIPOLA |
| 37 | ESTU_COD_DEPTO_PRESENTACION | Código del departamento de aplicación |
| 38 | ESTU_COD_OTRO_PAIS_PLANTEL | Código Dane del país donde estudio en el exterior |
| 39 | ESTU_COD_PLANTEL | Código del plantel educativo donde estudio |
| 40 | ESTU_COD_RESIDE_DEPTO | Código dane del departamento de residencia del inscrito |
| 41 | ESTU_COD_RESIDE_MCPIO | Código el municipio de residencia del estudiante |
| 42 | ESTU_CONSECUTIVO | Código de identificación de la persona en la base de datos de investigación |
| 43 | ESTU_CUANTO_FALTO_COLEGIO | Cantidad de veces que faltó al colegio |
| 44 | ESTU_CUANTOS_COLE_ESTUDIO | En cuantos colegios estudio |
| 45 | ESTU_DEPTO_PRESENTACION | Departamento de aplicación del examen |
| 46 | ESTU_EDAD | Edad del inscrito |
| 47 | ESTU_ESTUDIANTE | Identifica la condición académica en la que presento el examen |
| 48 | ESTU_ETNIA | Etnia a la que pertenece el inscrito |
| 49 | ESTU_FECHA_NACIMIENTO | Fecha de nacimiento del inscrito en formato DD/MM/AA |
| 50 | ESTU_EXAM_NOMBREEXAMEN | Nombre del examen presentado |
| 51 | ESTU_GENERO | Genero del inscrito |
| 52 | ESTU_HORAS_TRABAJA | Número de horas que trabaja a la semana |
| 53 | ESTU_IES_COD_DESEADA | SNP de la institución en la que desea estudiar la carrera |
| 54 | ESTU_IES_COD_MPIO_DESEADA | Código del municipio en el que se ubica la institución deseada |
| 55 | ESTU_IES_DEPT_DESEADA | Departamento en el que se ubica la institución deseada |
| 56 | ESTU_IES_DESEADA_NOMBRE | TOFIX Nombre de la institución en la que desea estudiar la carrera |
| 57 | ESTU_IES_MPIO_INS_DESEADA | Nombre de ciudad de la institución en la que desea estudiar la carrera |
| 58 | ESTU_IES_MPIO_DESEADA | Municipio en el que se ubica la institución deseada |
| 59 | ESTU_INGRESO_CARRERA | Ingreso carrera |
| 60 | ESTU_LIMITA_AUTISMO | Indicador de discapacidad autismo |
| 61 | ESTU_LIMITA_BAJAVISION | Indicador de discapacidad baja visión |
| 62 | ESTU_LIMITA_COGNITIVA | Indicador de discapacidad cognitiva |
| 63 | ESTU_LIMITA_CONDICIONESPECIAL | Indicador de condición especial |
| 64 | ESTU_LIMITA_INVIDENTE | Indicador de discapacidad invidente |
| 65 | ESTU_LIMITA_MOTRIZ | Indicador de discapacidad motriz |
| 66 | ESTU_LIMITA_SDOWN | Indicador de discapacidad |
| 67 | ESTU_LIMITA_SORDO | Indicador de discapacidad sordo |
| 68 | ESTU_LIMITA_SORDOCEGUERA | Indicador de discapacidad sordo ceguera |
| 69 | ESTU_LIMITA_SORDOINTERPRETE | Indicador de discapacidad sordo y requiere interprete de señas |

Tabla 13. (Continuación)

| | | |
|----|--------------------------------|--|
| 70 | ESTU_LIMITA_SORDONINTERPRETE | Indicador de discapacidad sordo y no requiere interprete de señas |
| 71 | ESTU_MCPIO_PRESENTACION | Municipio de aplicación del examen |
| 72 | ESTU_NUMERO_LIBROS | Cantidad de libros que existen en el hogar |
| 73 | ESTU_OTRO_PLANTEL | Colegios que ha estudiado |
| 74 | ESTU_MES_TERMINO_BACHILL | Mes en el que el estudiante termino el bachillerato. Solo aplica para INDIVIDUALES |
| 75 | ESTU_NACIMIENTO_ANNO | Año del nacimiento del inscrito |
| 76 | ESTU_NACIMIENTO_DIA | Día del nacimiento del inscrito |
| 77 | ESTU_NACIMIENTO_MES | Mes del nacimiento del inscrito |
| 78 | ESTU_PAIS_RESIDE | Código del país de residencia del inscrito con codificación ISO 3166 alpha2 |
| 79 | ESTU_POR_AMIGOSESTUDIANDO | Cuáles son las principales razones por las que desea estudiar en esa institución Amigos estudian en esa institución |
| 80 | ESTU_POR_BUSCANDOCARRERA | Opciones que influyeron en la elección del programa académico: la participación en uno de los talleres de encuentros juveniles del programa Buscando carrera del Ministerio de Educación nacional |
| 81 | ESTU_POR_COLOMBIAAPRENDE | Opciones que influyeron en la elección del programa académico: La información o guía buscando carrera contenida en el portal www.colombiaprende.edu.co o en otros |
| 82 | ESTU_POR_COSTOMATRICULA | Cuáles son las principales razones por las que desea estudiar en esa institución, El costo de la matrícula está a su alcance |
| 83 | ESTU_POR_INFLUENCIAALGUIEN | Influencia de amigos, padres o familiares |
| 84 | ESTU_POR_INTERESPERSONAL | Interés personal |
| 85 | ESTU_POR_MEJORARPOSICIONSOCIAL | Posibilidad de mejorar la posición social y/o económica |
| 86 | ESTU_POR_OPORTUNIDADES | Cuáles son las principales razones por las que desea estudiar en esa institución, Brinda mejores/mayores oportunidades de empleo, valor de la respuesta afirmativo, negativo |
| 87 | ESTU_POR_ORIENTACIONVOCACIONAL | Opciones que influyeron en la elección del programa académico: La orientación vocacional/profesional recibida en el colegio |
| 88 | ESTU_POR_OTRARAZON | Cuáles son las principales razones por las que desea estudiar en esa institución |
| 89 | ESTU_POR_TRADICIONFAMILIAR | Tradición familiar |
| 90 | ESTU_POR_UBICACION | Cuáles son las principales razones por las que desea estudiar en esa institución Por la ubicación y cercanía |
| 91 | ESTU_POR_UNICAQUEOFRECE | Cuáles son las principales razones por las que desea estudiar en esa institución, Es la única que ofrece la carrera que desea |
| 92 | ESTU_PRESENTO_ANTECEDENTES | Sin Descripción |
| 93 | ESTU_PRESENTO_ESPECTATIVAS | Expectativas del estudiante |

Tabla 13. (Continuación)

| | | |
|-----|--------------------------------|--|
| 94 | ESTU_PRESTIGIOINSTITUCION | Cuáles son las principales razones por las que desea estudiar en esa institución, Por el prestigio de la institución |
| 95 | ESTU_PRIVADO_LIBERTAD | Respuesta del evaluado si actualmente se encuentra privado de la libertad |
| 96 | ESTU_PROGRAMA_ACADEMICODESEADO | Tipo de programa académico deseado |
| 97 | ESTU_PROMEDIO_SALON | Promedio del estudiante en el salón |
| 98 | ESTU_PUESTO | Puesto general examen |
| 99 | ESTU_PUNT_ESPERADO_INGLES | Aspiración puntaje de ingles |
| 100 | ESTU_PUNT_ESPERADO LENGUAJE | Aspiración puntaje de lenguaje |
| 101 | ESTU_PUNT_ESPERADO_MATEMATICAS | Aspiración puntaje de matemáticas |
| 102 | ESTU_RAZON_RETIRO | Razones por las cuales se retiró de estudiar |
| 103 | ESTU_RECIBE_SALARIO | Recibe salario |
| 104 | ESTU_REPROBO_CUARTO | Grado de primaria que reprobó (perdió) o tuvo que repetir: Cuarto |
| 105 | ESTU_REPROBO_DECIMO | Grado de secundaria que reprobó (perdió) o tuvo que repetir: Decimo |
| 106 | ESTU_REPROBO_NOVENO | Grado de secundaria que reprobó (perdió) o tuvo que repetir: Noveno |
| 107 | ESTU_REPROBO_OCTAVO | Grado de secundaria que reprobó (perdió) o tuvo que repetir: Octavo |
| 108 | ESTU_REPROBO_ONCE_MAS | Grado de secundaria que reprobó (perdió) o tuvo que repetir: Undécimo o más |
| 109 | ESTU_REPROBO_PRIMERO | Grado de primaria que reprobó (perdió) o tuvo que repetir: Primero |
| 110 | ESTU_REPROBO_QUINTO | Grado de primaria que reprobó (perdió) o tuvo que repetir: Quinto |
| 111 | ESTU_REPROBO_SEGUNDO | Grado de primaria que reprobó (perdió) o tuvo que repetir: Segundo |
| 112 | ESTU_REPROBO_SEPTIMO | Grado de secundaria que reprobó (perdió) o tuvo que repetir: Séptimo |
| 113 | ESTU_REPROBO_SEXTO | Grado de secundaria que reprobó (perdió) o tuvo que repetir: Sexto |
| 114 | ESTU_REPROBO_TERCERO | Grado de primaria que reprobó (perdió) o tuvo que repetir: Tercero |
| 115 | ESTU_RESIDE_DEPTO | Departamento de residencia del inscrito |
| 116 | ESTU_RESIDE_MCPIO | Municipio de residencia del inscrito |
| 117 | ESTU_RETIRARSE_COLEGIO | Alguna vez tuvo que retirarse del colegio (suspender estudio) |
| 118 | ESTU_SALARIO_ESPERADO_BACH | Aspiración de salario como bachiller |
| 119 | ESTU_SALARIO_ESPERADO_PROFES | Aspiración de salario como profesional |
| 120 | ESTU_SALARIO_ESPERADO_TECNICO | Aspiración de salario como técnico |
| 121 | ESTU_TIPO_CARRERA_DESEADA | Qué tipo de carrera desea estudiar cuando obtenga el título de bachiller |
| 122 | ESTU_TIPO_DOCUMENTO | Tipo de documento de identidad del inscrito |
| 123 | ESTU_TOTAL_ALUMNOS_CURSO | Cuántos alumnos hay en total en el curso (salón) |
| 124 | ESTU TRABAJA | Trabaja actualmente |
| 125 | ESTU_VECES_ESTADO_ESTUDIANTIL | Veces que el estudiante ha presentado el examen |
| 126 | ESTU_VECES_ESTADO | Número de veces que ha presentado el examen de estado solo para ESTUDIANTES |
| 127 | ESTU_ZONA_RESIDE | Zona de residencia del inscrito |

Tabla 13. (Continuación)

| | | |
|-----|-------------------------------|---|
| 128 | FAMI_AUTOMOVIL | Bienes que posee su hogar: Automóvil particular |
| 129 | FAMI_CELULAR | Cantidad de celulares con que cuenta su hogar |
| 130 | FAMI_COMPUTADOR | Hogar hay computador |
| 131 | FAMI_CUARTOS_HOGAR | Número de habitaciones de la residencia |
| 132 | FAMI_DVD | Bienes que posee su hogar: Reproductor de DVD |
| 133 | FAMI_EDUCA_MADRE | Máximo nivel educativo alcanzada por la madre |
| 134 | FAMI_EDUCA_PADRE | Máximo nivel educativo alcanzada por el padre |
| 135 | FAMI_ESTRATO_VIVIENDA | Estrato socioeconómico de la residencia del estudiante según factura de energía |
| 136 | FAMI_HORNO | Hogar cuenta con horno |
| 137 | FAMI_INGRESO_FMILIAR_MENSUAL | Ingresos mensuales representado en salarios mínimos mensuales |
| 138 | FAMI_INTERNET | Hogar cuenta con conexión a internet |
| 139 | FAMI_LAVADORA | Hogar cuenta con lavadora |
| 140 | FAMI_MICROONDAS | Hogar cuenta con microondas |
| 141 | FAMI_NEVERA | Hogar cuenta con nevera |
| 142 | FAMI_NIVEL_SISBEN | Niel del Sisben en el que está clasificada la familia |
| 143 | FAMI_NUM_HERMANOS | Número de hermanos que tiene el inscrito |
| 144 | FAMI_OCUPA_MADRE | Ocupación de la madre |
| 145 | FAMI_OCUPA_PADRE | Ocupación del padre |
| 146 | FAMI_PERSONAS_HOGAR | Número de personas que forman el hogar |
| 147 | FAMI_PISOSHOGAR | Material de los pisos del hogar |
| 148 | FAMI_SERVICIO_TELEVISION | Hogar cuenta con servicio de televisión |
| 149 | FAMI_TELEFONO_FIJO | Hogar cuenta con teléfono fijo |
| 150 | PERCENTIL_C_NATURALES | Percentil ciencias naturales |
| 151 | PERCENTIL_GLOBAL | Percentil total del evaluado |
| 152 | PERCENTIL_INGLES | Percentil en inglés |
| 153 | PERCENTIL_LECTURA_CRITICA | Percentil lectura crítica |
| 154 | PERCENTIL_MATEMATICAS | Percentil en matemáticas |
| 155 | PERCENTIL_SICIALES_CIUDADANAS | Percentil sociales y ciudadanas |
| 156 | FAMI_TELEVISOR | Servicios con los que cuenta en su hogar: Televisión |
| 157 | IND_COD_ICFES_TERMINO | Código icfes |
| 158 | NOMBRE_COMP_FLEXIBLE | Nombre del componente flexible (profundización o interdisciplinarias) |
| 159 | PERIODO | Periodo en el que se aplicó el examen |
| 160 | PUNT_BIOLOGIA | Puntaje |
| 161 | PUNT_C_NATURALES | Puntaje |
| 162 | PUNT_C_SOCIALES | Puntaje |
| 163 | PUNT_COMP_CIDADANA | Puntaje |
| 164 | PUNT_COMP_FLEXIBLE | Puntaje |
| 165 | PUNT_FILOSOFIA | Puntaje |
| 166 | PUNT_FISICA | Puntaje |
| 167 | PUNT_GLOBAL | Puntaje |
| 168 | PUNT_INGLES | Puntaje |
| 169 | PUNT_LECTURA_CRITICA | Puntaje |
| 170 | PUNT LENGUAJE | Puntaje |
| 171 | PUNT_MATEMATICAS | Puntaje |
| 172 | PUNT_QUIMICA | Puntaje |
| 173 | PUNT_RAZONA_CUANT | Puntaje |
| 174 | PUNT_SOCIALES_CIUDADANAS | Puntaje |

Fuente: ICFES acceso a bases de datos

Continuando con el análisis de los atributos, en las Tablas 14 y 15 se presenta la comparación de los atributos para identificar los más importantes para el desarrollo del mercado de datos.

Tabla 14. Descripción colores tabla 15.

| COLOR | DESCRIPCIÓN |
|-------|------------------------------|
| | No incluida en el periodo |
| | Aparece en todas las pruebas |
| | Aparece en 9/10 pruebas |
| | Aparece en 8/10 pruebas |
| | Aparece en 7/10 pruebas |
| | Aparece en 6/10 pruebas |
| | Aparece en 5/10 pruebas |
| | Aparece en 4/10 pruebas |
| | Aparece en 3/10 pruebas |
| | Aparece en 2/10 pruebas |
| | Aparece en 1/10 pruebas |

Tabla 15. Atributos de los repositorios fuentes.

| No | VARIABLES | 2012-1 | 2012-2 | 2013-1 | 2013-2 | 2014-1 | 2014-2 | 2015-1 | 2015-2 | 2016-1 | 2016-2 |
|----|---------------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 1 | COLE_AREA_UBICACION | | | | | | x | x | x | x | X |
| 2 | COLE_BILINGUE | x | x | x | x | | x | x | x | | |
| 3 | COLE_CALENDARIO | x | x | x | x | | x | x | x | x | x |
| 4 | COLE_CARACTER | x | x | x | x | | x | x | x | x | x |
| 5 | COLE_COD_DANE_INSTITUCION | x | x | x | x | x | | | x | x | x |
| 6 | COLE_COD_ICFES | x | x | x | x | x | x | x | x | x | x |
| 7 | COLE_COD_MCPIO_UBICACION | | | | | | x | x | x | x | x |
| 8 | COLE_DEPTO_UBICACION | | | | | | x | x | x | | |
| 9 | COLE_GENERO | x | x | x | x | | x | x | x | x | x |
| 10 | COLE_JORNADA | x | x | x | x | x | x | x | x | x | x |
| 11 | COLE_MCPIO_UBICACION | | | | | | x | x | x | | |
| 12 | COLE_NATURALEZA | x | x | x | x | | x | x | x | x | x |
| 13 | COLE_NOMBRE_SEDE | x | x | | x | x | x | x | x | x | x |
| 14 | COLE_VALOR_PENSION | x | x | x | x | x | x | x | x | x | x |
| 15 | DECIL_C_NATURALES | | | | | | | | x | | |

| | | | | | | | | | | | | | |
|-----|--------------------------------|---|---|---|---|---|---|---|---|---|---|---|---|
| 98 | ESTU_PUESTO | x | x | x | x | x | x | x | x | | | | |
| 99 | ESTU_PUNT_ESPERADO_INGLES | | | | | | | | x | | | | |
| 100 | ESTU_PUNT_ESPERADO LENGUAJE | | | | | | | | x | | | | |
| 101 | ESTU_PUNT_ESPERADO_MATEMATICAS | | | | | | | | x | | | | |
| 102 | ESTU_RAZON_RETIRO | x | x | x | x | x | x | x | | | | | |
| 103 | ESTU_RECIBE_SALARIO | | | | | | x | x | x | x | x | x | |
| 104 | ESTU_REPROBO_CUARTO | x | x | x | x | x | x | x | | | | | |
| 105 | ESTU_REPROBO_DECIMO | x | x | x | x | x | x | x | | | | | |
| 106 | ESTU_REPROBO_NOVENO | x | x | x | x | x | x | x | | | | | |
| 107 | ESTU_REPROBO_OCTAVO | x | x | x | x | x | x | x | | | | | |
| 108 | ESTU_REPROBO_ONCE_MAS | x | x | x | x | x | x | x | | | | | |
| 109 | ESTU_REPROBO_PRIMERO | x | x | x | x | x | x | x | | | | | |
| 110 | ESTU_REPROBO_QUINTO | x | x | x | x | x | x | x | | | | | |
| 111 | ESTU_REPROBO_SEGUNDO | x | x | x | x | x | x | x | | | | | |
| 112 | ESTU_REPROBO_SEPTIMO | x | x | x | x | x | x | x | | | | | |
| 113 | ESTU_REPROBO_SEXTO | x | x | x | x | x | x | x | | | | | |
| 114 | ESTU_REPROBO_TERCERO | x | x | x | x | x | x | x | | | | | |
| 115 | ESTU_RESIDE_DEPTO | x | x | x | x | x | x | x | x | x | x | x | |
| 116 | ESTU_RESIDE_MCPIO | x | | x | x | x | x | x | x | x | x | x | |
| 117 | ESTU_RETIRARSE_COLEGIO | x | x | x | x | x | x | x | | | | | |
| 118 | ESTU_SALARIO_ESPERADO_BACH | | | | | | | | x | | | | |
| 119 | ESTU_SALARIO_ESPERADO_PROFES | | | | | | | | x | | | | |
| 120 | ESTU_SALARIO_ESPERADO_TECNICO | | | | | | | | x | | | | |
| 121 | ESTU_TIPO_CARRERA_DESEADA | | | | | | | | x | x | | | |
| 122 | ESTU_TIPO_DOCUMENTO | x | x | x | x | | | x | x | x | x | x | |
| 123 | ESTU_TOTAL_ALUMNOS_CURSO | x | x | x | x | x | x | x | | | | | |
| 124 | ESTU TRABAJA | x | x | x | x | x | x | x | x | x | x | x | |
| 125 | ESTU_VECES_ESTADO_ESTUDIANTIL | | | | | | | | x | x | x | x | x |
| 126 | ESTU_VECES_ESTADO | x | x | x | x | x | x | x | x | x | x | x | |
| 127 | ESTU_ZONA_RESIDE | x | x | x | x | x | x | x | x | x | x | x | |
| 128 | FAMI_AUTOMOVIL | x | x | x | x | x | x | x | x | x | x | x | |
| 129 | FAMI_CELULAR | x | x | x | x | x | x | | | | | | |
| 130 | FAMI_COMPUTADOR | x | x | | x | x | x | x | x | x | x | x | |
| 131 | FAMI_CUARTOS_HOGAR | x | x | x | x | x | x | x | x | x | x | x | |
| 132 | FAMI_DVD | x | x | x | x | x | x | x | | | | | |
| 133 | FAMI_EDUCA_MADRE | x | x | x | x | x | x | x | x | x | x | x | |
| 134 | FAMI_EDUCA_PADRE | x | x | x | x | x | x | x | x | x | x | x | |
| 135 | FAMI_ESTRATO_VIVIENDA | x | x | x | x | x | x | x | x | x | x | x | |
| 136 | FAMI_HORNO | x | x | x | x | x | x | x | x | x | x | x | |
| 137 | FAMI_INGRESO_FMILIAR_MENSUAL | x | x | x | x | x | x | x | x | x | x | x | |
| 138 | FAMI_INTERNET | x | x | x | x | x | x | x | x | x | x | x | |
| 139 | FAMI_LAVADORA | x | x | x | x | x | x | x | x | x | x | x | |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|-----|-------------------------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 140 | FAMI_MICROONDAS | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | | | | | | | | | | | | | | | | | | | | |
| 141 | FAMI_NEVERA | x | x | x | x | x | x | | | | | | | | x | x | x | | | | | | | | | | | | | | | | | | | | | | |
| 142 | FAMI_NIVEL_SISBEN | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | | | | | | | | | | | | | | | | | | | | |
| 143 | FAMI_NUM_HERMANOS | | | | | | | | | | | | | | x | | | x | x | | | | | | | | | | | | | | | | | | | | |
| 144 | FAMI_OCUPA_MADRE | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | | | | | | | | | | | | | | | | | | | | |
| 145 | FAMI_OCUPA_PADRE | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | | | | | | | | | | | | | | | | | | | | |
| 146 | FAMI_PERSONAS_HOGAR | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | | | | | | | | | | | | | | | | | | | | |
| 147 | FAMI_PISOSHOGAR | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | | | | | | | | | | | | | | | | | | | | |
| 148 | FAMI_SERVICIO_TELEVISION | x | x | x | x | | | | | x | | | | | | | x | x | x | | | | | | | | | | | | | | | | | | | | |
| 149 | FAMI_TELEFONO_FIJO | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | | | | | | | | | | | | | | | | | | | | |
| 150 | PERCENTIL_C_NATURALES | | | | | | | | | | | | | | | | | | x | x | | | | | | | | | | | | | | | | | | | |
| 151 | PERCENTIL_GLOBAL | | | | | | | | | | | | | | | | | | x | x | | | | | | | | | | | | | | | | | | | |
| 152 | PERCENTIL_INGLES | | | | | | | | | | | | | | | | | | x | x | | | | | | | | | | | | | | | | | | | |
| 153 | PERCENTIL_LECTURA_CRITICA | | | | | | | | | | | | | | | | | | x | x | | | | | | | | | | | | | | | | | | | |
| 154 | PERCENTIL_MATEMATICAS | | | | | | | | | | | | | | | | | | x | x | | | | | | | | | | | | | | | | | | | |
| 155 | PERCENTIL_SICIALES_CIUDADANAS | | | | | | | | | | | | | | | | | | x | x | | | | | | | | | | | | | | | | | | | |
| 156 | FAMI_TELEVISOR | | | | | | | | | x | | | | | | | x | | | | | | | | | | | | | | | | | | | | | | |
| 157 | IND_COD_ICFES_TERMINO | x | x | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 158 | NOMBRE_COMP_FLEXIBLE | x | x | x | x | x | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 159 | PERIODO | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | | | | | | | | | | | | | | | | | | | |
| 160 | PUNT_BIOLOGIA | x | x | x | x | x | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 161 | PUNT_C_NATURALES | | | | | | | | | | | | | | | | | | | x | x | x | x | x | | | | | | | | | | | | | | | |
| 162 | PUNT_C_SOCIALES | x | x | x | x | x | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 163 | PUNT_COMP_CIDADANA | | | | | | | | | | | | | | | | | | | | | | | x | x | x | | | | | | | | | | | | | |
| 164 | PUNT_COMP_FLEXIBLE | x | x | x | x | x | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 165 | PUNT_FILOSOFIA | x | x | x | x | x | x | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 166 | PUNT_FISICA | x | x | x | x | x | x | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 167 | PUNT_GLOBAL | | | | | | | | | | | | | | | | | | | | | | | | | x | x | x | x | x | | | | | | | | | |
| 168 | PUNT_INGLES | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | | | | | | | | | | |
| 169 | PUNT_LECTURA_CRITICA | | | | | | | | | | | | | | | | | | | | | | | | | | | | | x | x | x | x | x | | | | | |
| 170 | PUNT LENGUAJE | x | x | x | x | x | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 171 | PUNT_MATEMATICAS | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | | | | | | | |
| 172 | PUNT_QUIMICA | x | x | x | x | x | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 173 | PUNT_RAZONA_CUANT | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | x | x | x | | | |
| 174 | PUNT_SOCIALES_CIUDADANAS | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | x | x | x | x | x |

Análisis de la calidad de datos

A partir del repositorio saber11_2012_2016 se creó el repositorio saber11_2012_2016_limpio tomando como estándar la tabla saber11_2016_2 con el propósito de utilizar los datos necesarios y realizar un análisis inicial de la calidad de los datos e identificar el número de registros no nulo, nulo, distintos valores y valores distintos. Este repositorio cuenta con 81 atributos y 3'356.299 registros.

En la tabla 16 se muestran los resultados del análisis.

Tabla 14. Análisis de calidad de datos.

| No. | Atributo | # Nulos | # No Nulos | # Distintos | Distintos |
|-----|---------------------------|---------|------------|-------------|--|
| 1 | cole_area_ubicacion | 1648095 | 1708204 | 2 | "R" "U" |
| 2 | cole_calendario | 351484 | 3004815 | 4 | "F" "B" "A" "O" |
| 3 | cole_caracter | 370730 | 2985569 | 9 | "TÉCNICO/ACADÉMICO" "ACADÉMICO" "ACADEMICO" "TÉCNICO" "TECNICO" "NORMALISTA" "ACADEMICO Y TECNICO" "NO_APLICA" "DESCONOCIDO" |
| 4 | cole_cod_dane_institucion | 297267 | 3059032 | 11787 | "000000000000" "015232300063" "016340100097" "016354800229" "022524500284" "027021501099" "027610900000" "031100192201" "031100192996" "031383800202" |
| 5 | cole_cod_icfes | 279261 | 3077038 | 16289 | "000002" "000003" "000018" "000034" "000042" "000059" "000067" "000075" "000083" "000091" |

Tabla 16. (Continuación)

| | | | | | |
|----|--------------------------|---------|---------|-------|--|
| 6 | cole_cod_mcpio_ubicacion | 1648078 | 1708221 | 1113 | "05001" "05002" "05004" "05021" "05030" "05031" "05034" "05036" "05038" "05040" |
| 7 | cole_genero | 359012 | 2997287 | 6 | "F" "FEMENINO" "M" "MASCULINO" "MIXTO" "X" |
| 8 | cole_jornada | 287413 | 3068886 | 12 | "C" U "COMPLETA ORDINARIA" "M" "MAÑANA" "N" "NOCHE" "S" "SABATINA DOMINICAL" - "T" "TARDE" "U" "UNICA" |
| 9 | cole_naturaleza | 131346 | 3224953 | 5 | "N" OFICIAL" "NO OFICIAL" "O" "OFICIAL" "OTRO" |
| 10 | cole_nombre_sede | 279261 | 3077038 | 14361 | "07/01 INSTITUCION EDUCATIVA JORGE ISAACS INEM" "20/01 INSTITUCION EDUCATIVA, SIETE DE AGOSTO" "28/01 INSTITUCIÓN EDUCATIVA SANTA FE" "32/01 INSTITUCION EDUCATIVA NORMAL SUPERIOR SANTIAGO DE CALI" |

Tabla 16. (Continuación)

| | | | | | |
|----|-----------------------------|---------|---------|-----|--|
| 11 | cole_valor_pension | 481497 | 2874802 | 13 | "0" "1" "10" "11" "12" "2" "3" "4" "5" "6" |
| 12 | desemp_ingles | 691 | 3355608 | 5 | "A-" "A1" "A2" "B+ " "B1" |
| 13 | estu_ano_termino_bachill | 3136160 | 220139 | 38 | "1980" "1981" "1982" "1983" "1984" "1985" "1986" "1987" "1988" "1989" |
| 14 | estu_area_reside | 680437 | 2675862 | 4 | "1" "2" "R" "U" |
| 15 | estu_cod_depto_presentacion | 1420409 | 1935890 | 33 | "05" "08" "11" "13" "15" "17" "18" "19" "20" "23" "25" |
| 16 | estu_cod_mcpio_presentacion | 4 | 3356295 | 468 | "05001" "05002" "05030" "05031" "05034" "05042" "05045" "05051" "05079" "05088" |

Tabla 16. (Continuación)

| | | | | | |
|----|-------------------------|---------|---------|---------|--|
| 17 | estu_cod_reside_depto | 1424408 | 1931891 | 34 | "11" "13" "15" "17" "18" "19" "20" "23" "25" "27" |
| 18 | estu_cod_reside_mcpio | 7657 | 3348642 | 1119 | "05001" "05002" "05004" "05021" "05030" "05031" "05034" "05036" "05038" "05040" |
| 19 | estu_consecutivo | 0 | 3356299 | 3356299 | "SABER112012100 0001" "SABER112012100 0002" "SABER112012100 0003" "SABER112012100 0004" "SABER112012100 0005" "SABER112012100 0006" "SABER112012100 0007" "SABER112012100 0008" "SABER112012100 0009" "SABER112012100 0010" "SABER112012100 0011" |
| 20 | estu_depto_presentacion | 4 | 3356295 | 44 | "AMAZONAS" "ANTIOQUIA" "ARAUCA" "ATLANTICO" "ATLÁNTICO" "BOGOTA" "BOGOTÁ" "BOLIVAR" "BOLÍVAR" "BOYACA" |

Tabla 16. (Continuación)

| | | | | | |
|----|-------------------------------|---------|---------|-------|--|
| 21 | estu_edad | 88784 | 3267515 | 119 | "0" "1" "10" "100" "106" "11" "113" "115" "117" "12" |
| 22 | estu_estudiante | 1420409 | 1935890 | 2 | "ESTUDIANTE" "INDIVIDUAL" |
| 23 | estu_etnia | 2467485 | 888814 | 30 | "01" "02" "03" "04" "05" "06" "07" "08" "09" "1" |
| 24 | estu_fecha_nacimiento | 1332947 | 2023352 | 29573 | "00/00/0000" "01/01/00" "01/01/1900" "01/01/1942" "01/01/1950" "01/01/1953" "01/01/1955" "01/01/1956" "01/01/1957" "01/01/1958" |
| 25 | estu_genero | 10440 | 3345859 | 2 | "F" "M" |
| 26 | estu_limita_autismo | 3266766 | 89533 | 2 | "-" "AA" |
| 27 | estu_limita_condicionespecial | 3266273 | 90026 | 2 | "-" "CE" |
| 28 | estu_limita_invidente | 3355429 | 870 | 2 | "I" "VIS" |
| 29 | estu_limita_motriz | 3352442 | 3857 | 2 | "M" "MOT" |
| 30 | estu_limita_sdown | 3267172 | 89127 | 2 | "-" "SD" |
| 31 | estu_limita_sordo | 3267170 | 89129 | 2 | "-" "AUD" |
| 32 | estu_mcpio_presentacion | 4 | 3356295 | 452 | "ABEJORRAL" "ABREGO" "ACACIAS" "ACANDI" "ACHI" "AGUACHICA" "AGUADAS" "AGUAZUL" "AGUSTIN CODAZZI" "AIPE" |

Tabla 16. (Continuación)

| | | | | | |
|----|--------------------------|---------|---------|-----|--|
| 33 | estu_mes_termino_bachill | 3135482 | 220817 | 12 | "1" "10" "11" "12" "2" "3" "4" "5" "6" "7" |
| 34 | estu_pais_reside | 0 | 3356299 | 129 | "1" "10" "101" "103" "104" "GT" "HN" "IL" "IN" "IR" |
| 35 | estu_privado_libertad | 1439312 | 1916987 | 2 | "N" "S" |
| 36 | estu_recibe_salario | 3108687 | 247612 | 3 | "-" "N" "S" |
| 37 | estu_reprobo_octavo | 3231411 | 124888 | 4 | "0" "1" "N" "S" |
| 38 | estu_reprobo_once_mas | 3231411 | 124888 | 4 | "0" "1" "N" "S" |
| 39 | estu_reprobo_primero | 3231411 | 124888 | 4 | "0" "1" "N" "S" |
| 40 | estu_reprobo_quinto | 3231411 | 124888 | 4 | "0" "1" "N" "S" |
| 41 | estu_reprobo_segundo | 3231412 | 124887 | 4 | "0" "1" "N" "S" |
| 42 | estu_reprobo_septimo | 3231411 | 124888 | 4 | "0" "1" "N" "S" |
| 43 | estu_reprobo_sexto | 3231412 | 124887 | 4 | "0" "1" "N" "S" |

Tabla 16. (Continuación)

| | | | | | |
|----|-------------------------------|---------|---------|------|---|
| 44 | estu_reprobo_tercero | 3231411 | 124888 | 4 | "0" "1" "N" "S" |
| 45 | estu_reside_depto | 7658 | 3348641 | 45 | "AMAZONAS" "ANTIOQUIA" "ARAUCA" "ATLANTICO" "ATLÁNTICO" "BOGOTA" "BOGOTÁ" "BOLIVAR" "BOLÍVAR" "BOYACA" |
| 46 | estu_reside_mcpio | 7657 | 3348642 | 1038 | "ABEJORRAL" "ABREGO" "ABRIAQUI" "ACACIAS" "ACANDI" "ACEVEDO" "ACHI" "AGRADO" "AGUACHICA" "AGUADA" |
| 47 | estu_retirarse_colegio | 1873218 | 1483081 | 5 | "-" "1" "2" "3" "4" |
| 48 | estu_salario_esperado_bach | 3300802 | 55497 | 6 | "1" "2" "3" "4" "5" "6" |
| 49 | estu_salario_esperado_profes | 3300802 | 55497 | 6 | "1" "2" "3" "4" "5" "6" |
| 50 | estu_salario_esperado_tecnico | 3300802 | 55497 | 6 | "1" "2" "3" "4" "5" "6" |
| 51 | estu_tipo_carrera_deseada | 3293627 | 62672 | 4 | "1" "2" "3" "4" |

Tabla 16. (Continuación)

| | | | | | |
|----|-------------------------------|---------|---------|-----|---|
| 52 | estu_tipo_documento | 88739 | 3267560 | 17 | "C" "CC" "CE" "CR" "E" "NIP" "NUI" "P" "PC" "PE" |
| 53 | estu_total_alumnos_curso | 3169483 | 186816 | 100 | "0" "1" "10" "11" "12" "13" "14" "15" "16" "17" |
| 54 | estu_trabaja | 463028 | 2893271 | 8 | "0" "1" "2" "3" "4" "5" "6" "7" |
| 55 | estu_veces_estado | 1779030 | 1577269 | 5 | "0" "1" "2" "3" "4" |
| 56 | estu_veces_estado_estudiantil | 218515 | 3137784 | 9 | "0" "1" "2" "3" "4" "CUATRO" "DOS" "TRES" "UNA" |
| 57 | estu_zona_reside | 3841 | 3352458 | 10 | "1" "10" "2" "3" "4" "5" "6" "7" "8" "9" |
| 58 | fami_automovil | 20035 | 3336264 | 4 | "0" "1" "N" "S" |

Tabla 16. (Continuación)

| | | | | | |
|----|-----------------------|---------|---------|----|--|
| 59 | fami_celular | 1368909 | 1987390 | 4 | "0" "1" "N" "S" |
| 60 | fami_computador | 20033 | 3336266 | 5 | "0" "1" "3" "N" "S" |
| 61 | fami_cuartos_hogar | 666400 | 2689899 | 10 | "1" "10" "2" "3" "4" "5" "6" "7" "8" "9" |
| 62 | fami_dvd | 20034 | 3336265 | 4 | "0" "1" "N" "S" |
| 63 | fami_educa_madre | 20091 | 3336208 | 19 | "0" "1" "10" "11" "12" "13" "14" "15" "16" "17" |
| 64 | fami_educa_padre | 20091 | 3336208 | 19 | "0" "1" "10" "11" "12" "13" "14" "15" "16" "17" |
| 65 | fami_estrato_vivienda | 650909 | 2705390 | 8 | "0" "1" "2" "3" "4" "5" "6" "8" |
| 66 | fami_horno | 20036 | 3336263 | 4 | "0" "1" "N" "S" |

Tabla 16. (Continuación)

| | | | | | |
|----|------------------------------|---------|---------|----|---|
| 67 | fami_ingreso_fmiliar_mensual | 20090 | 3336209 | 7 | "1" "2" "3" "4" "5" "6" "7" |
| 68 | fami_internet | 20034 | 3336265 | 4 | "0" "1" "N" "S" |
| 69 | fami_lavadora | 20035 | 3336264 | 4 | "0" "1" "N" "S" |
| 70 | fami_microondas | 20035 | 3336264 | 4 | "0" "1" "N" "S" |
| 71 | fami_nevera | 1368908 | 1987391 | 4 | "0" "1" "N" "S" |
| 72 | fami_nivel_sisben | 20093 | 3336206 | 5 | "1" "2" "3" "4" "5" |
| 73 | fami_ocupa_madre | 20093 | 3336206 | 33 | "01" "02" "03" "04" "05" "06" "07" "08" "09" "1" |
| 74 | fami_ocupa_padre | 20092 | 3336207 | 34 | "01" "02" "03" "04" "05" "06" "07" "08" "09" "1" |

Tabla 16. (Continuación)

| | | | | | |
|----|--------------------------|---------|---------|----|--|
| 75 | fami_personas_hogar | 20125 | 3336174 | 12 | "1" "10" "11" "12" "2" "3" "4" "5" "6" "7" |
| 76 | fami_pisoshogar | 20091 | 3336208 | 5 | "1" "2" "3" "4" "5" |
| 77 | fami_servicio_television | 216001 | 3140298 | 4 | "0" "1" "N" "S" |
| 78 | fami_telefono_fijo | 20035 | 3336264 | 4 | "0" "1" "N" "S" |
| 79 | periodo | 0 | 3356299 | 10 | "20121" "20122" "20131" "20132" "20141" "20142" "20151" "20152" "20161" "20162" |
| 80 | punt_c_naturales | 1420409 | 1935890 | 94 | "0" "1" "10" "100" "13" "14" "15" "16" "17" "18" "19" |
| 81 | punt_comp_ciudadana | 2100518 | 1255781 | 80 | "0" "10" "100" "11" "12" "14" "15" "16" "17" "18" |

- Limpieza de datos

Se nombran algunos de los procedimientos que se ejecutaron sobre el repositorio saber11_2012_2016_limpio con el propósito de reemplazar valores nulos, corregir valores erróneos, eliminación de datos innecesarios y en general mejorar las condiciones de calidad del repositorio:

- Creación repositorio saber11_2012_2016_limpio a partir de saber11_2012_2016 (Anexo 1).
- Reemplazar el valor del atributo estu_reside_depto de “NARIÑO” por NARINO. (Anexo 2).
- Reemplazar el valor del atributo cole_jornada de “MAÑANA” por el valor MANANA, “SABATINA-DOMINICAL” por SABATINA_DOMINICAL (Anexo 4).
- Reemplazar el valor ‘U’ del atributo cole_jornada por ‘UNICA’ (Anexo 5).
- Reemplazar el valor ‘S’ del atributo cole_jornada por ‘SABATINA DOMINICAL’ (Anexo 6).
- Reemplazar el valor ‘C’ del atributo cole_jornada por ‘COMPLETA U ORDINARIA’ (Anexo 7).
- Reemplazar el valor ‘M’ del atributo cole_jornada por ‘MANANA’ (Anexo 8).
- Reemplazar el valor ‘T’ del atributo cole_jornada por ‘TARDE’ (Anexo 9).
- Reemplazar el valor ‘N’ del atributo cole_jornada por ‘NOCHE’ (Anexo 10).
- Reemplazar los valores del atributo cole_caracter de “ACADÉMICO” por ACADEMICO, “TÉCNICO” por TECNICO, “TÉCNICO/ACADÉMICO” por ACADEMICO Y TECNICO (Anexo 11).
- Reemplazar el valor ‘1’ del atributo estu_area_reside por ‘U’ (Anexo 12).
- Reemplazar el valor ‘2’ del atributo estu_area_reside por ‘R’ (Anexo 13).
- Reemplazar el valor ‘F’ del atributo cole_genero por ‘FEMENINO’ (Anexo 14).

- Reemplazar el valor 'M' del atributo cole_genero por 'MASCULINO' (Anexo 15).
- Reemplazar el valor 'X' del atributo cole_genero por 'MIXTO' (Anexo 16).
- Reemplazar el valor 'O' del atributo cole_naturaleza por 'OFICIAL' (Anexo 17).
- Reemplazar el valor 'N' del atributo cole_naturaleza por 'NO OFICIAL' (Anexo 18).
- Reemplazar el valor 'N' del atributo FAMI_DVD por '0' (Anexo 19).
- Reemplazar el valor 'S' del atributo FAMI_DVD por '1' (Anexo 20).
- Reemplazar el valor 'N' del atributo FAMI_TELEFONO_FIJO por '0' (Anexo 21)
- Reemplazar los valores 'S' del atributo FAMI_TELEFONO_FIJO por '1' (Anexo 22)
- Reemplazar los valores 'N' del atributo FAMI_TELEVISOR por '0' (Anexo 23).
- Reemplazar los valores 'S' del atributo FAMI_TELEVISOR por '1' (Anexo 24).
- Reemplazar los valores 'N' del atributo FAMI_INTERNET por '0' (Anexo 25).
- Reemplazar los valores 'S' del atributo FAMI_AUTOMOVIL por '1' (Anexo 26).

Con el proceso de limpieza se obtiene un repositorio con mejor calidad. Este cuenta con 81 atributos y 18.119 registros específicamente de la subregión de Obando.

1.3.2 Fase de transformación. El objetivo de esta fase fue construir un repositorio limpio y transformado con datos de calidad el cual fue utilizado como fuente de carga del mercado de datos. Para tal fin se hizo transformación y eliminación de datos.

Por consiguiente, se nombran algunos de los procesos que se ejecutaron sobre el repositorio saber11_2012_2016_transformado.

- **Transformación de Datos**

A partir del repositorio saber11_2012_2016_limpio se creó el repositorio saber11_2012_2016_transformado, sobre el cual se ejecutaron procedimientos para normalizar el valor de los atributos y facilitar posteriormente el análisis de resultados. Se nombran algunos de los procedimientos realizados.

- Creación repositorio saber11_2012_2016_transformado a partir de saber11_2012_2016_limpio (Anexo 29).
- Creación del atributo estu_edad_presentacion de tipo text (Anexo 30).
- Actualización del atributo estu_edad_presentacion (Anexo 31).
- Creación del atributo estu_rango_edad de tipo text (Anexo 32).
- Actualización del atributo estu_rango_edad (Anexo 33).
- Creación atributo estu_discapacidad de tipo text (Anexo 34).
- Creación del atributo fami_cod_max_nivel_educa_padres (Anexos 35).
- Actualización del atributo fami_cod_max_nivel_educa_padres con el máximo nivel educativo entre el padre y la madre (Anexos 36).
- Creación del atributo fami_max_nivel_educa_padres de tipo text (Anexos 37).
- Actualización del atributo fami_max_nivel_educa_padres con la descripción del código fami_cod_max_nivel_educa_padres. (Anexos 38).
- Creación del atributo fami_ocup_padre de tipo text (Anexo 39).
- Creación del atributo fami_ocup_padre de tipo text (Anexo 40).
- Actualización de los atributos fami_ocup_madre y fami_ocup_padre según código de los atributos fami_ocupa_padre y fami_ocupa_madre (Anexo 41).
- Reemplazar los códigos del atributo fami_nivel_sisben (Anexo 42).

- Creación del atributo eco_condicion_vive (Anexo 43).
- Actualización del atributo eco_condicion_vive de acuerdo a los valores del atributo fami_cuartos_hogar y fami_personas_hogar según índice de hacinamiento (Anexo 44).
- Creación del atributo eco_condicion_vivienda tipo text. (Anexo 45).
- Actualización del atributo eco_condicion_vivienda de acuerdo a la discretización del atributo fami_pisoshogar de la siguiente tabla (Anexo 46).
- Creación del atributo eco_condicion_tic (Anexo 47).
- Actualización del atributo eco_condicion_tic teniendo en cuenta los valores de los atributos fami_telefono_fijo, fami_celular, fami_internet, fami_servicio_television, fami_computador que forman este índice de acuerdo a la sumatoria de los valores de la siguiente tabla (Anexo 48).
- Creación del atributo eco_condicion_electrodomesticos (Anexo 49).
- Actualización del atributo eco_condicion_electrodomesticos teniendo en cuenta los valores de los atributos fami_nevera, fami_horno, fami_microondas, fami_lavadora, fami_dvd (Anexo 50).
- Convertir los valores del atributo fami_ingreso_fmiliar_mensual de códigos a descripción según la siguiente tabla (Anexo 51).
- Creación del atributo pun_global (Anexo 52).
- Actualización del atributo pun_global a partir de la suma de los puntajes de los atributos punt_biologia, punt_c_sociales, punt_comp_flexible, punt_filosofia, punt_fisica, punt_ingles, punt_lenguaje, punt_matematicas, punt_quica (Anexo 53).
- Creación del atributo prom_rend (Anexo 54).
- Actualización del atributo prom_rend a partir del atributo $(\text{pun_global}) / 9$ y $(\text{pun_global}) / 5$ (Anexo 55).
- Creación del atributo rendimiento (Anexo 56).

- Actualización del atributo rendimiento a partir del atributo prom_rend (Anexo 57).

- **Eliminación de Datos**

Utilizando el repositorio saber11_2012_2016_transformado se creó el repositorio saber11_2012_2016_final, sobre el cual se ejecutaron procedimientos para eliminar atributos que no son requeridos en el mercado de datos. Finalmente se obtuvo un repositorio con 42 atributos.

Teniendo en cuenta lo anterior, se nombran algunos de los procedimientos que se ejecutaron sobre el repositorio saber11_2012_2016_final.

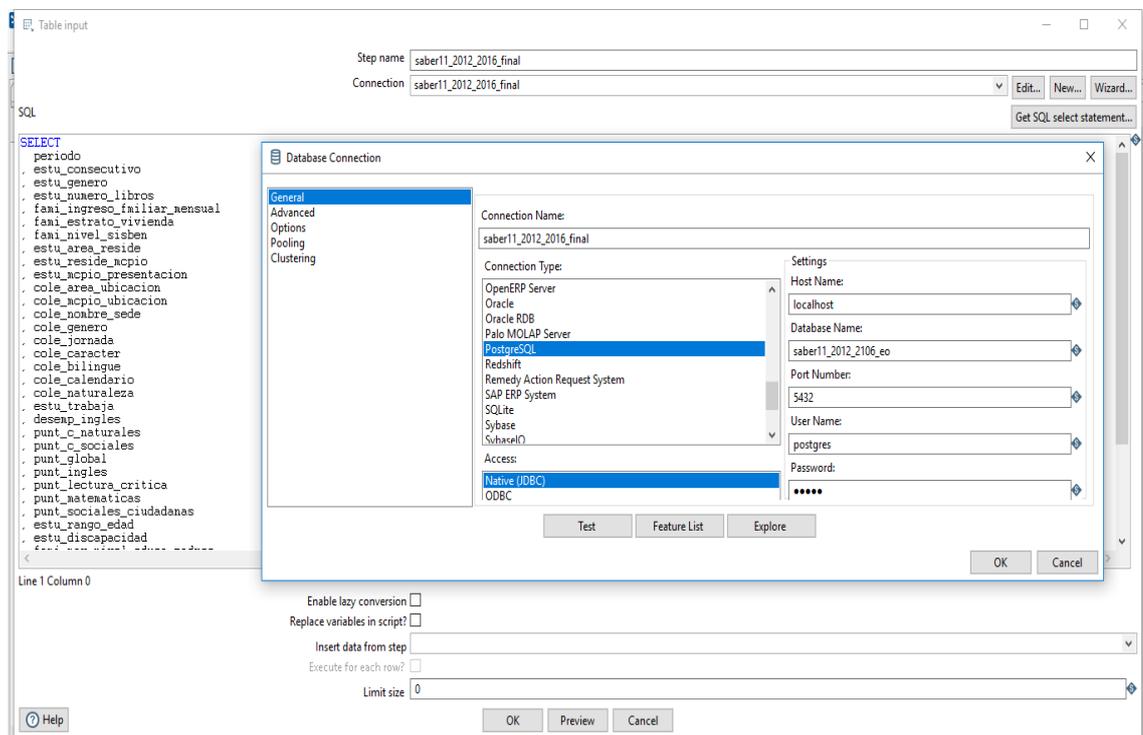
- Creación de repositorio saber11_2012_2016_final a partir de saber11_2012_2016_tranformado (anexo 58)
- Eliminación del atributo DECIL_C_NATURALES (Anexo 59).
- Eliminación del atributo DECIL_COMP_CIUADADANA (Anexo 60).
- Eliminación del atributo DECIL_INGLES (Anexo 61).
- Eliminación del atributo DECIL_LECTURA_CRITICA (Anexo 62).
- Eliminación del atributo DECIL_MATEMATICAS (Anexo 63).
- Eliminación del atributo DECIL_RAZONA_CUANT (Anexo 64).
- Eliminación del atributo DECIL_SOCIALES_CIUADADANAS (Anexo 65).
- Eliminación del atributo ESTU_CUANTO_FALTO_COLEGIO (Anexo 66).
- Eliminación del atributo DECIL_C_NATURALES (Anexo 67).
- Eliminación del atributo DESEMP_LECTURA_CRITICA (Anexo 68).
- Eliminación del atributo DESEMP_MATEMATICAS (Anexo 69).
- Eliminación del atributo DESEMP_SOCIALES_CIUADADANAS (Anexo 70).
- Eliminación del atributo PERCENTIL_C_NATURALES (Anexo 71).
- Eliminación del atributo PERCENTIL_GLOBAL (Anexo 72).
- Eliminación del atributo PERCENTIL_INGLES (Anexo 73).

- Eliminación del atributo PERCENTIL_LECTURA_CRITICA (Anexo 74).
- Eliminación del atributo PERCENTIL_MATEMATICAS (Anexo 75).
- Eliminación del atributo PERCENTIL_SICIALES_CIUDADANAS (Anexo 76).
- Eliminación del atributo FAMI_TELEVISOR (anexo 77)

1.3.3 Fase de carga. A partir del repositorio saber11_2012_2016_final se creó con la estructura de un mercado de datos, la base de datos final llamada Datamart, la cual contiene 9 dimensiones y 1 tabla de hechos. El paso siguiente fue cargar los datos con la herramienta PDI (Pentaho Data Integration) al mercado de datos.

En la figura 8 se observa como previamente antes de cargar los datos a las diferentes tablas del mercado de datos, se hizo la configuración de parámetros de conexión con el repositorio principal (saber11_2012_2016_final).

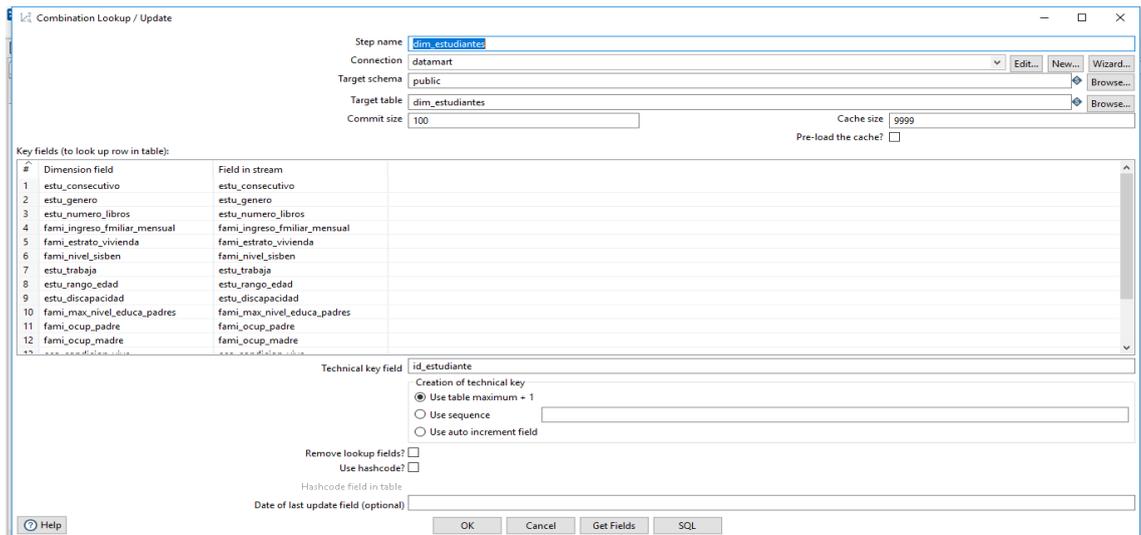
Figura 7. Configuración de parámetros.



Fuente: Esta investigación

Posteriormente a la conexión con el repositorio principal (saber11_2012_2016_final) se procede a cargar los datos a las diferentes dimensiones y tabla de hechos que conforman el mercado de datos. En la figura 9 se observa la configuración de parámetros de conexión para proceder a cargar los datos a la dimensión estudiante.

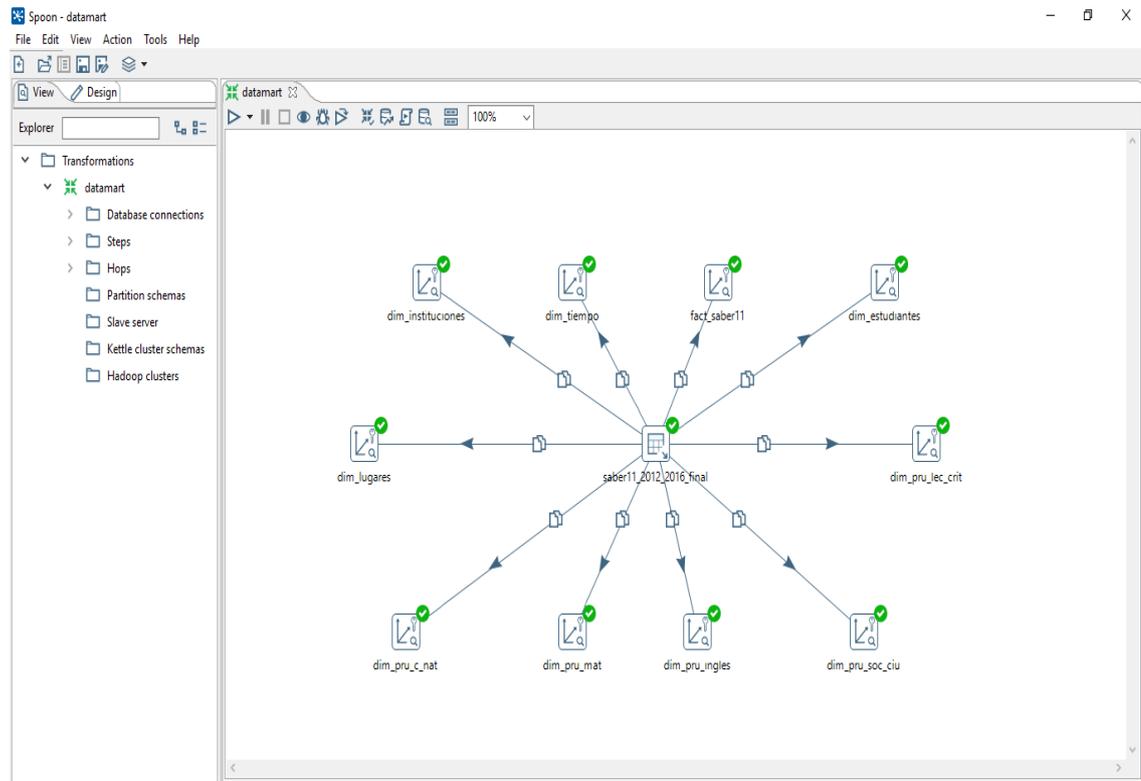
Figura 8. Proceso carga de datos dimensión estudiante.



Fuente: Esta investigación

Después de realizar los procesos requeridos de configuración para todas las dimensiones y tabla de hechos; en la figura 10 se presenta el proceso de carga de datos automatizado a partir del repositorio principal saber11_2012_2016_limpio.

Figura 9. Proceso carga de datos automatizado.

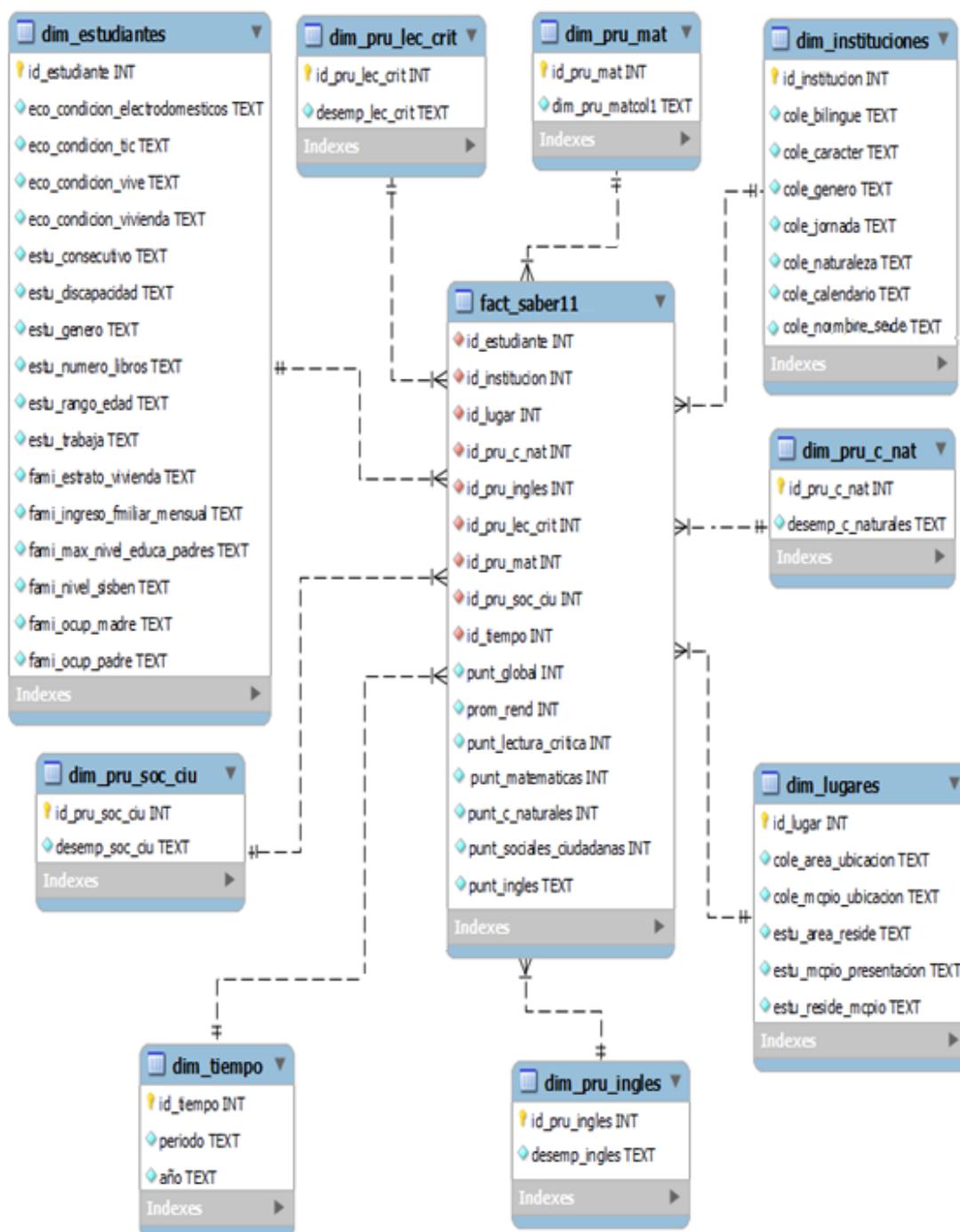


Fuente: Esta investigación

1.4 Implementación del mercado de datos en PostgreSQL

En la figura 11 se percibe la Implementación del Modelo Multidimensional tipo estrella que se encuentra en PostgreSQL el cual está conformado por 9 dimensiones cada una con sus respectivas llaves primarias y variables relevantes; Y una tabla de hechos conformada por las diferentes llaves foráneas y variables relevantes.

Figura 10. Modelo Multidimensional tipo estrella en PostgreSQL.



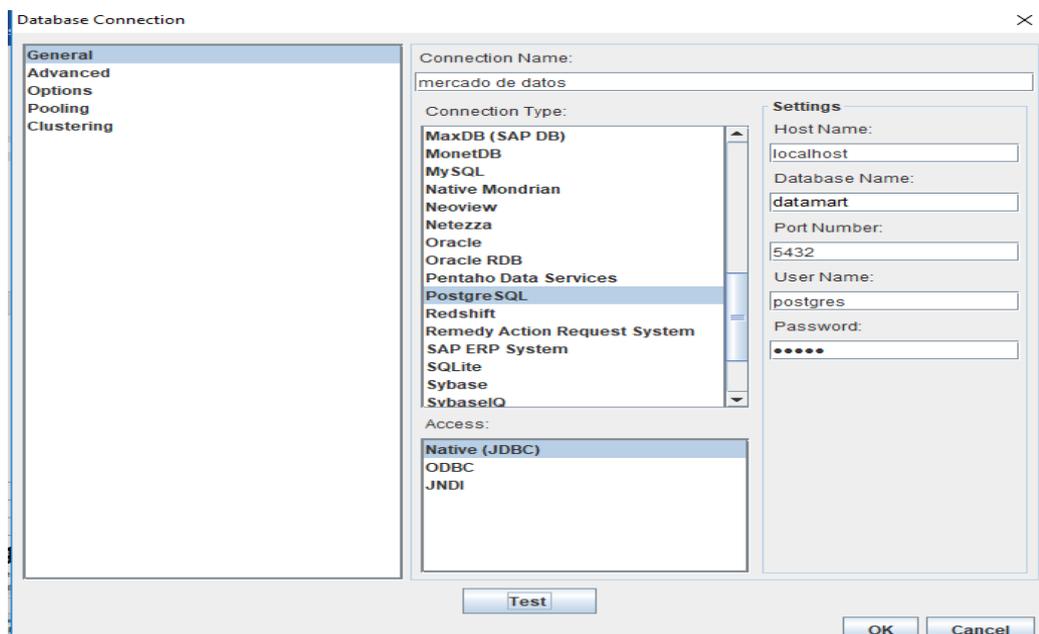
Fuente: Esta investigación

1.5 Integración del mercado de datos con Pentaho BI

1.5.1 **Cubo de datos.** A continuación, se presenta la configuración del cubo para consultas OLAP, realizada en la aplicación Schema Workbench de Mondrian.

Como primer paso se configuró los parámetros de conexión con el mercado de datos como se observa en la figura 12.

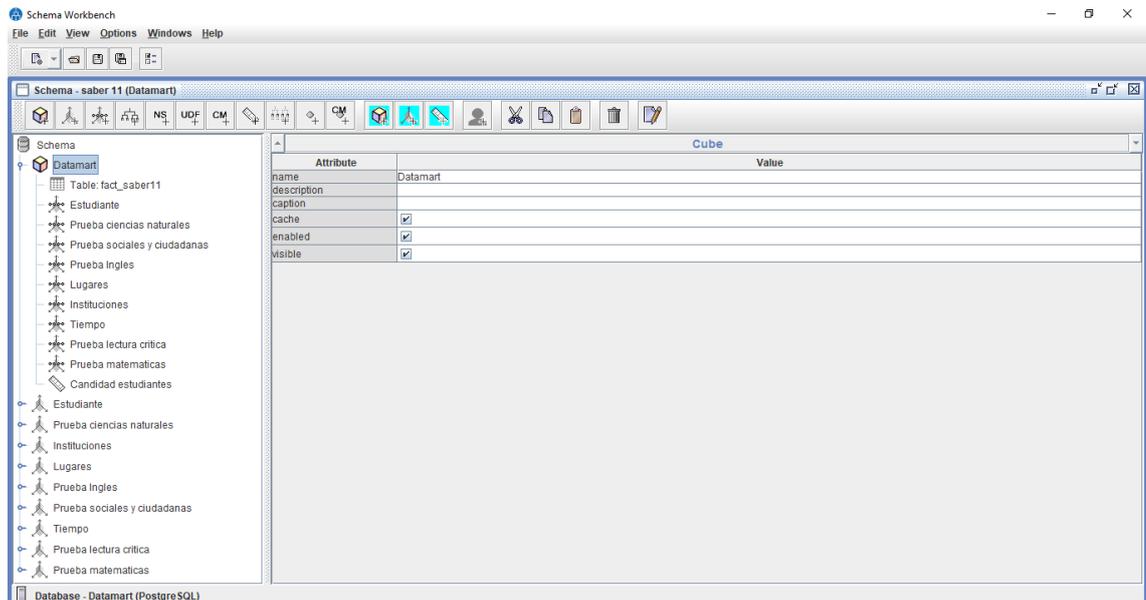
Figura 11. Parámetros de conexión



Fuente: Esta investigación

Después de la configuración de parámetros de conexión se procede a la creación del Schema, en el cual como primer paso se creó las diferentes dimensiones las cuales facilitan la creación del cubo, ya que para la creación del cubo solo se necesita asociar la tabla de hechos con las dimensiones necesarias ya creadas y una unidad de medida. Como se observa en la figura 13.

Figura 12. Creación de cubo en Schema Workbench.



Fuente: Esta investigación

1.5.2 Operaciones OLAP:

A continuación, en la tabla 17 se presenta las operaciones OLAP roll up y slice&dice dónde se observa el total de estudiantes que presentaron la prueba en los diferentes periodos.

Tabla 15. Operación OLAP roll up y slice&dice

| Período | 2012_1 | 2012_2 | 2013_1 | 2013_2 | 2014_1 | 2014_2 | 2015_1 | 2015_2 | 2016_1 | 2016_2 |
|--------------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Total estudiantes | 819 | 3097 | 658 | 3220 | 792 | 3028 | 72 | 3201 | 83 | 3149 |

En la tabla 18 se presenta las operaciones olap drill down y slice&dice dónde se observa el total de estudiantes que presentaron la prueba en los diferentes municipios y periodos.

Tabla 16. Operación olap drill down y slice&dice

| Municipio del colegio | Periodo | | | | | | | | | |
|--------------------------|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | 2012_1 | 2012_2 | 2013_1 | 2013_2 | 2014_1 | 2014_2 | 2015_1 | 2015_2 | 2016_1 | 2016_2 |
| Aldana | 18 | 96 | 18 | 98 | 11 | 9696 | | 94 | | 91 |
| Contadero | 6 | 44 | 9 | 40 | 13 | 52 | | 60 | | 58 |
| Córdoba | 30 | 160 | 32 | 210 | 45 | 171 | | 201 | | 176 |
| Cuaspud Carlosama | 5 | 77 | 5 | 60 | 4 | 88 | | 69 | | 73 |

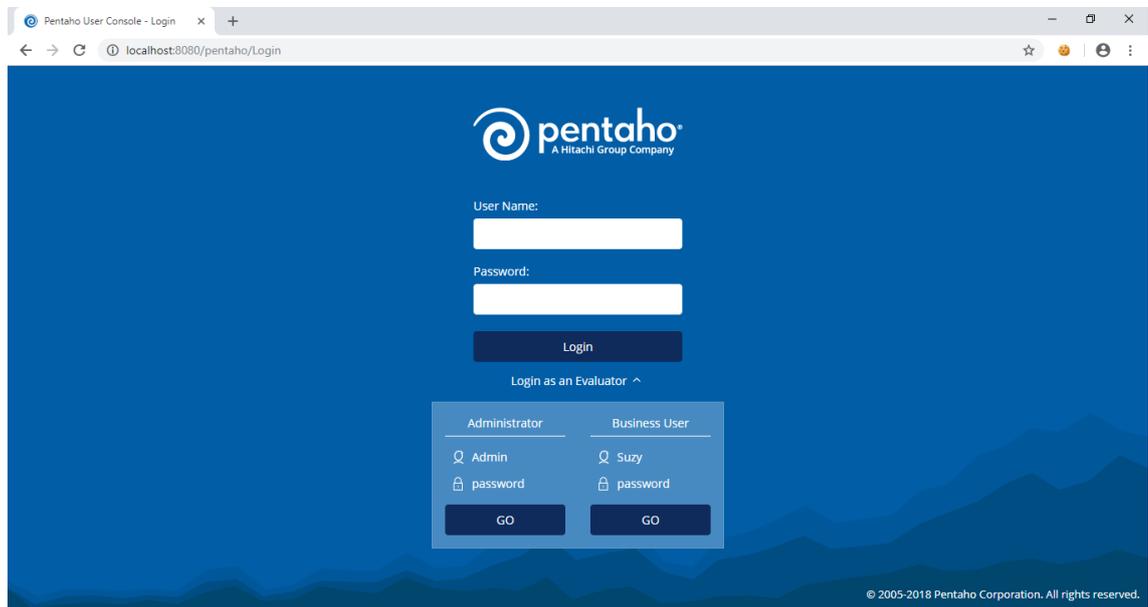
| | | | | | | | | | | |
|------------------|-----|------|-----|------|-----|------|----|------|----|------|
| Cumbal | 143 | 452 | 72 | 413 | 95 | 401 | | 446 | | 449 |
| Funes | 12 | 60 | 4 | 66 | 4 | 77 | | 79 | | 83 |
| Guachucal | 48 | 204 | 32 | 225 | 54 | 233 | | 194 | | 227 |
| Gualmatán | 6 | 38 | 1 | 36 | 13 | 91 | | 93 | | 77 |
| Iles | 28 | 134 | 16 | 118 | 23 | 110 | | 103 | | 87 |
| Ipiales | 420 | 1455 | 394 | 1549 | 444 | 1334 | 72 | 1526 | 83 | 1501 |
| Potosí | 2 | 59 | 4 | 61 | 3 | 72 | | 60 | | 59 |
| Puerres | 38 | 124 | 30 | 107 | 32 | 95 | | 79 | | 70 |
| Pupiales | 63 | 194 | 41 | 237 | 51 | 208 | | 197 | | 198 |

1.6 Especificación y desarrollo de aplicaciones de BI

Para el diseño e implementación de los cubos que representa cada una de las medidas sobre los hechos se utilizó el programa libre pentaho-server de la suite de Pentaho BI (Pentaho Community Edition).

Después de realizar los procesos requeridos para la configuración, se diseñó cada uno de los cubos y posteriormente mediante la suite de pentaho-server, se ajustaron las características de visualización del cubo. El procedimiento se muestra en las Figuras 14 a 16.

Figura 13. Pantalla de inicio de la suite de Pentaho.



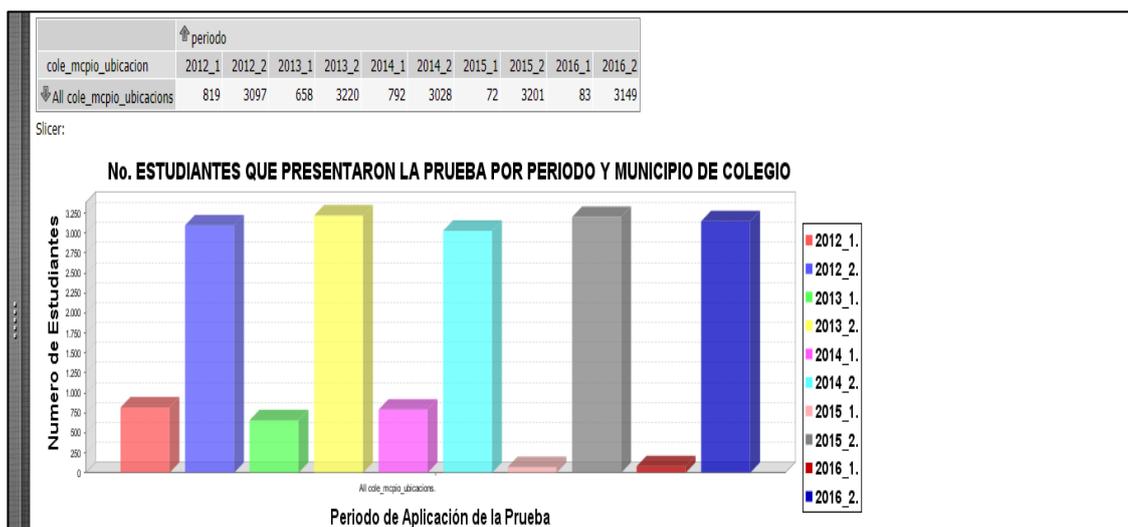
Fuente: Esta investigación

Figura 15. Visualización del cubo de datos.

| cole_mcpio_ubicacion | periodo | | | | | | | | | |
|----------------------|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | 2012_1 | 2012_2 | 2013_1 | 2013_2 | 2014_1 | 2014_2 | 2015_1 | 2015_2 | 2016_1 | 2016_2 |
| ALDANA | 18 | 96 | 18 | 98 | 11 | 96 | | 94 | | 91 |
| CONTADERO | 6 | 44 | 9 | 40 | 13 | 52 | | 60 | | 58 |
| CORDOBA | 30 | 160 | 32 | 210 | 45 | 171 | | 201 | | 176 |
| CUASPUD CARLOSAMA | 5 | 77 | 5 | 60 | 4 | 88 | | 69 | | 73 |
| CUMBAL | 143 | 452 | 72 | 413 | 95 | 401 | | 446 | | 449 |
| FUNES | 12 | 60 | 4 | 66 | 4 | 77 | | 79 | | 83 |
| GUACHUCAL | 48 | 204 | 32 | 225 | 54 | 233 | | 194 | | 227 |
| GUAIMATAN | 6 | 38 | 1 | 36 | 13 | 91 | | 93 | | 77 |
| ILES | 28 | 134 | 16 | 118 | 23 | 110 | | 103 | | 87 |
| IPIALES | 420 | 1455 | 394 | 1549 | 444 | 1334 | 72 | 1526 | 83 | 1501 |
| POTOSI | 2 | 59 | 4 | 61 | 3 | 72 | | 60 | | 59 |
| PUERRES | 38 | 124 | 30 | 107 | 32 | 95 | | 79 | | 70 |
| PUPIALES | 63 | 194 | 41 | 237 | 51 | 208 | | 197 | | 198 |

Fuente: Esta investigación

Figura 16. Gráfico asociado al cubo de datos.



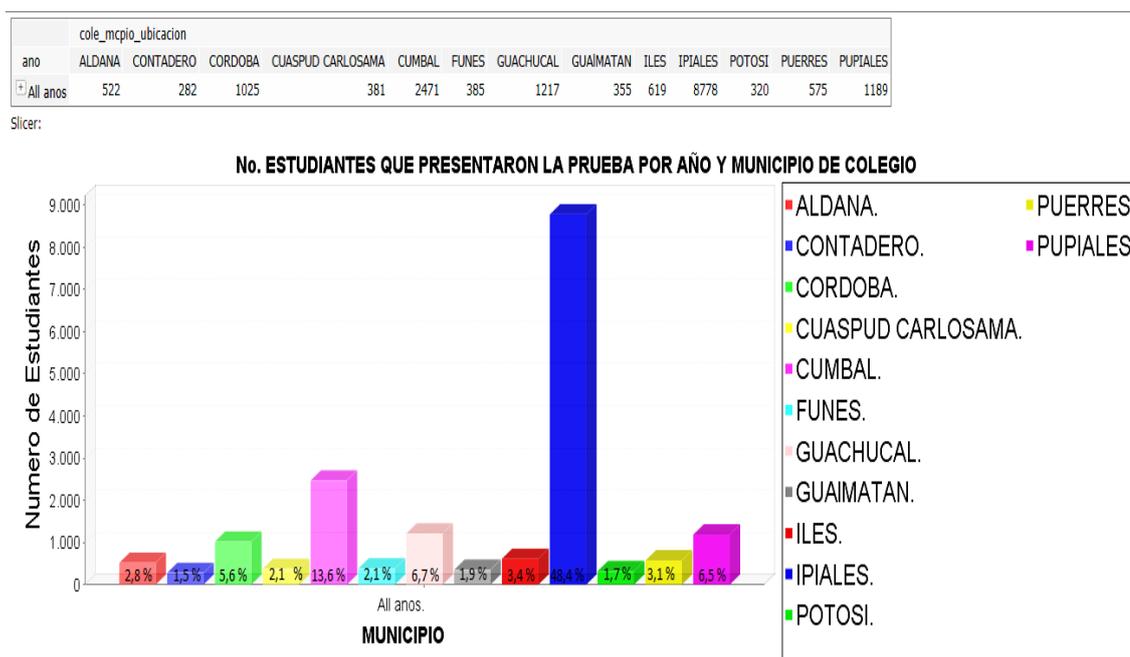
Fuente: Esta investigación

2 DISCUSIÓN Y RESULTADOS

A partir del modelo multidimensional se construyó diferentes cubos para representar cada una de las medidas sobre los hechos:

¿Cuántos estudiantes presentaron las Pruebas Saber 11° por municipio?

Figura 17. Número de estudiantes que presentaron la prueba por municipio.



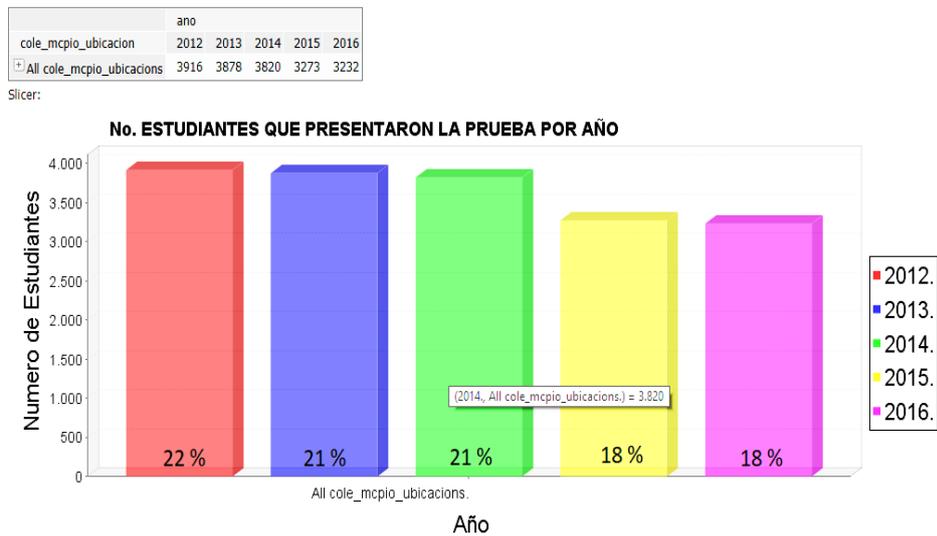
Fuente: Esta investigación

En la figura 17 se observa que en la Subregión de Obando del 2012 al 2016 el municipio que tiene mayor participación en estas pruebas es Ipiales con un 48%, seguido por Cumbal con el 13 % y Guachucal con el 6 %.

Por otra parte, el municipio con menor participación en estas pruebas es el Contadero con el 1.5 %, seguido de Potosí con el 1.7% y Gualmatán con el 1.9%.

- ¿Cuántos estudiantes presentaron las Prueba Saber 11° por año?

Figura 18. Número de estudiantes que presentaron la Prueba por año.



Fuente: Esta investigación

En la figura 18 se observa que en el año con más participación de estas pruebas en la subregión de Obando es el año 2012, y con menos participación es en el 2016.

También se puede concluir que en la Subregión de Obando cada año la población de estudiantes que presenta la Prueba es menor, afirmando así que del 2012 al 2016 esta población ha disminuido en un 4%.

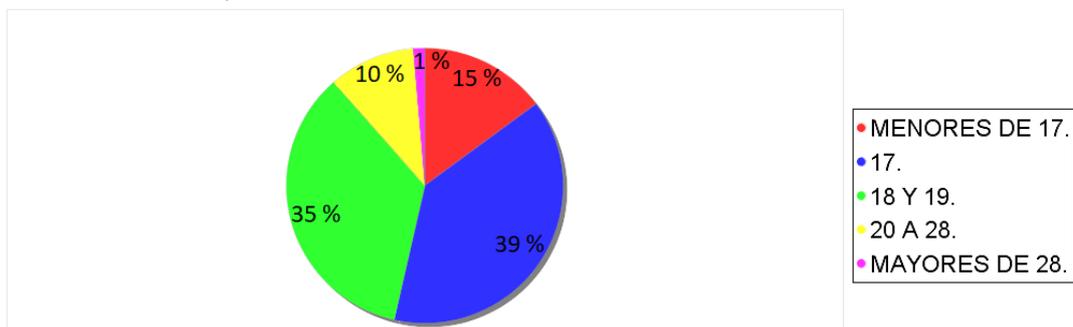
- ¿Cuántos estudiantes presentaron la Prueba Saber 11° por rango de edades?

Figura 19. Número de estudiantes por rangos de edades.

| desemp_soc_ciu | |
|-----------------|---------------------|
| estu_rango_edad | All desemp_soc_cius |
| MENORES DE 17 | 2696 |
| 17 | 7000 |
| 18 Y 19 | 6346 |
| 20 A 28 | 1833 |
| MAYORES DE 28 | 244 |

Slicer:

No. ESTUDIANTES QUE PRESENTARON LA PRUEBA DE SOCIALES CIUDADANAS POR RANGOS DE EDAD



Fuente: Esta investigación

En la figura 19 se observa que en la Subregión de Obando desde 2012 al 2016 la mayor parte de los estudiantes que presentan la prueba tienen 17 años, siendo esto equivalente al 39 %, seguido de los estudiantes que tienen 18 y 19 equivalentes al 35 % y los menores de 17 equivalentes al 15 %.

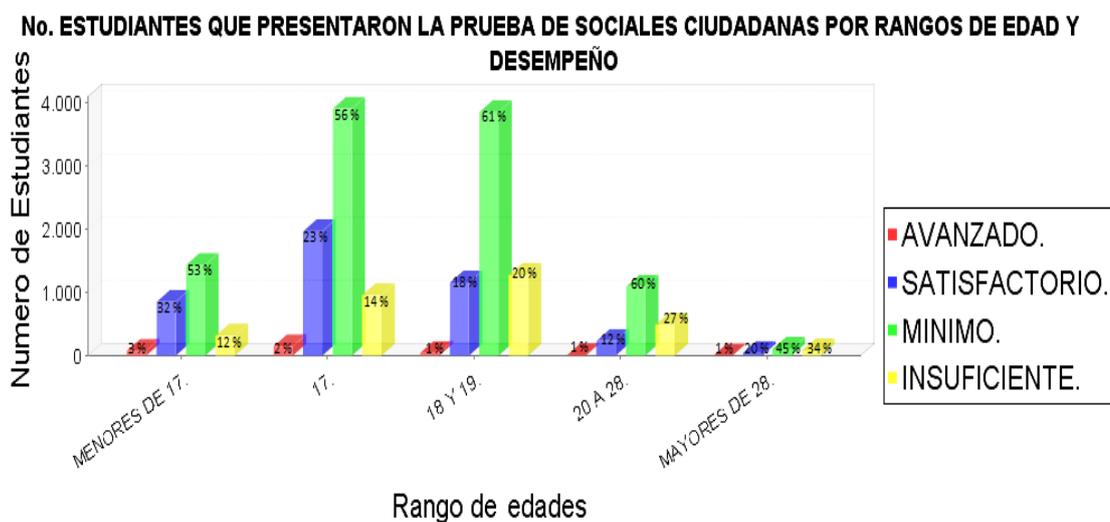
La menor población de los estudiantes que presentan la prueba son los mayores de 28 años equivalentes al 1 %.

- ¿Cuántos estudiantes presentaron la Prueba de sociales y ciudadanas por rango de edades y desempeño?

Figura 20. Número de estudiantes que presentaron la prueba de sociales y ciudadanas.

| estu_rango_edad | desemp_soc_ciu | | | |
|-----------------|----------------|---------------|--------|--------------|
| | AVANZADO | SATISFACTORIO | MINIMO | INSUFICIENTE |
| MENORES DE 17 | 79 | 855 | 1440 | 322 |
| 17 | 151 | 1973 | 3915 | 961 |
| 18 Y 19 | 47 | 1160 | 3862 | 1277 |
| 20 A 28 | 9 | 234 | 1096 | 494 |
| MAYORES DE 28 | 1 | 48 | 109 | 86 |

Slicer:



Fuente: Esta investigación

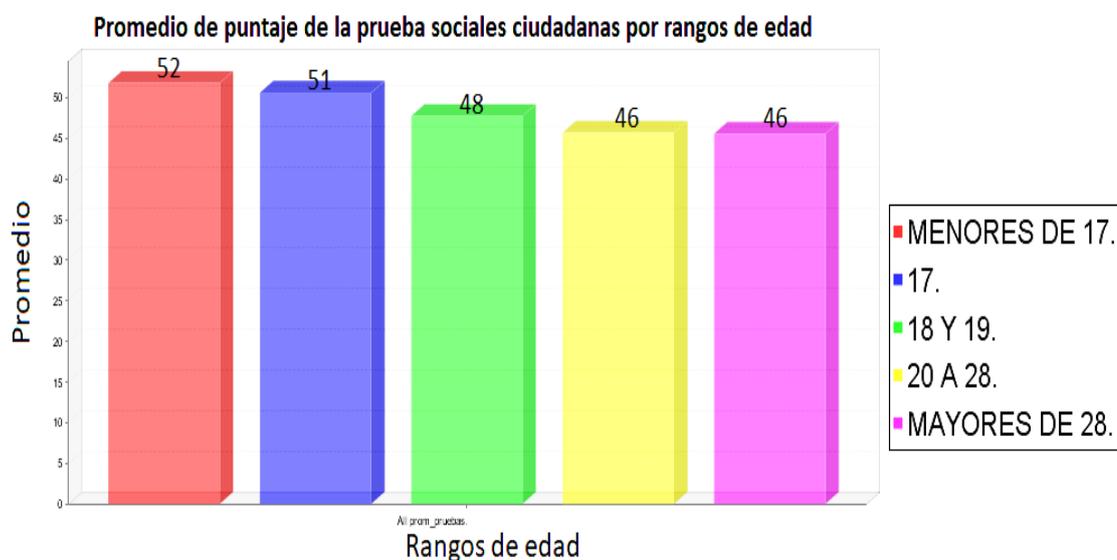
En la figura 20 se observa que los estudiantes menores de 18 años tienen un mejor rendimiento en el desempeño satisfactorio con respecto a las demás edades. El desempeño satisfactorio aumenta lo cual genera que el desempeño insuficiente disminuya.

- ¿Cuáles son los promedios de desempeño de la Prueba de sociales ciudadanas, ciencias naturales y matemáticas discriminados por rango de edad?

Figura 21. Promedios de desempeño de la Prueba de sociales y ciudadanas discriminado por rango de edad.

| | | estu_rango_edad | | | | |
|------------------|--|-----------------|----|---------|---------|---------------|
| prom_prueba | | MENORES DE 17 | 17 | 18 Y 19 | 20 A 28 | MAYORES DE 28 |
| All prom_pruebas | | 52 | 51 | 48 | 46 | 46 |

Slicer:



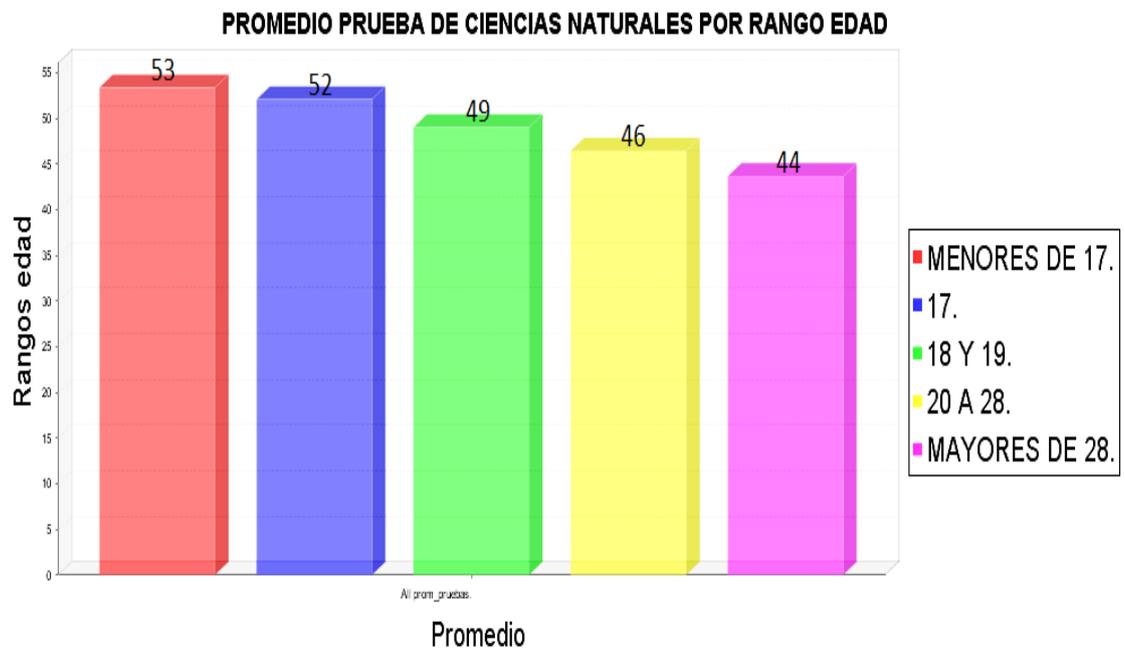
Fuente: Esta investigación

En la figura 21 se observa que los estudiantes con mejor promedio en la prueba de sociales ciudadanas son menores de 18 años.

Figura 22. Promedios de desempeño de la Prueba de ciencias naturales discriminado por rango de edad.

| prom_prueba | estu_rango_edad | | | | |
|------------------|-----------------|----|---------|---------|---------------|
| | MENORES DE 17 | 17 | 18 Y 19 | 20 A 28 | MAYORES DE 28 |
| All prom_pruebas | 53 | 52 | 49 | 46 | 44 |

Slicer:



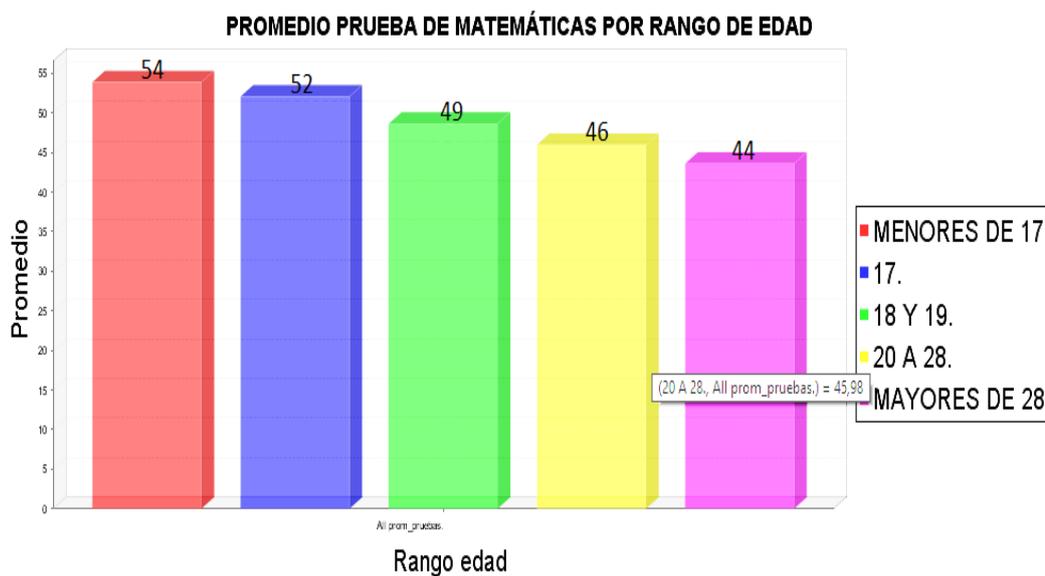
Fuente: Esta investigación

En la figura 22 se observa que los estudiantes con el menor promedio en la prueba de ciencias naturales son mayores de 28 años.

Figura 23. Promedios de desempeño de la Prueba de sociales y ciudadanas discriminado por rango de edad.

| | estu_rango_edad | | | | |
|------------------|-----------------|----|---------|---------|---------------|
| prom_prueba | MENORES DE 17 | 17 | 18 Y 19 | 20 A 28 | MAYORES DE 28 |
| All prom_pruebas | 54 | 52 | 49 | 46 | 44 |

Slicer:



Fuente: Esta investigación

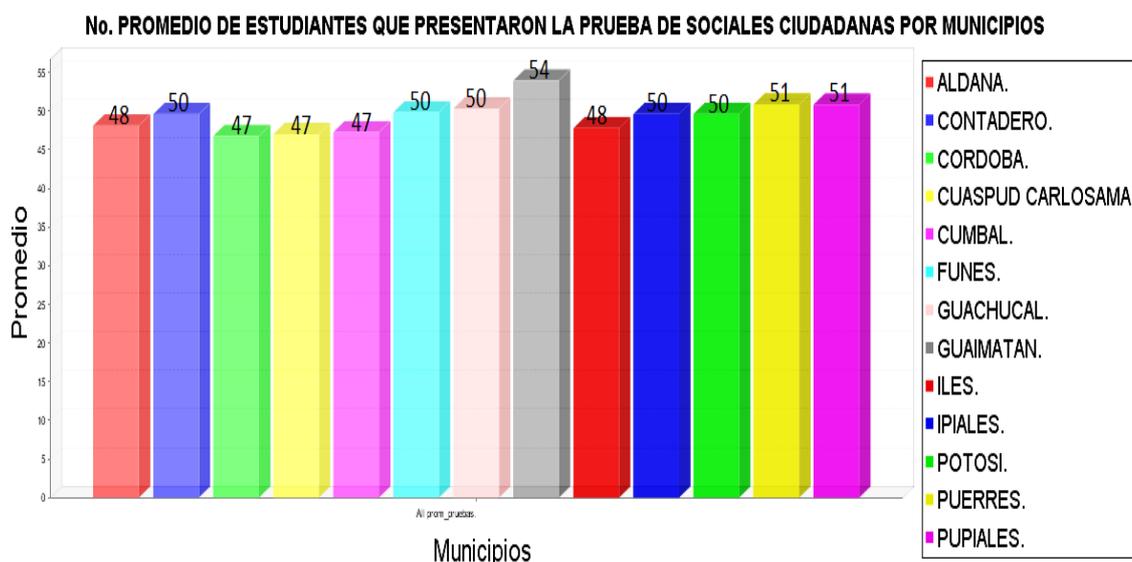
A pesar de que todos los rangos de edad tienen un bajo promedio en las pruebas de sociales ciudadanas, ciencias naturales y matemáticas, cabe resaltar que los estudiantes menores de 18 años tienen mejor promedio que las demás edades.

- ¿Cuáles son los promedios de desempeño de la Prueba de sociales ciudadanas, ciencias naturales y matemáticas describiendo por municipios?

Figura 24. Promedios de desempeño de la Prueba de sociales y ciudadanas discriminado por municipios.

| | cole_mcpio_ubicacion | | | | | | | | | | | | | |
|------------------|----------------------|-----------|---------|-------------------|--------|-------|-----------|-----------|------|---------|--------|---------|----------|--|
| prom_prueba | ALDANA | CONTADERO | CORDOBA | CUASPUD CARLOSAMA | CUMBAL | FUNES | GUACHUCAL | GUAIMATAN | ILES | IPIALES | POTOSI | PUERRES | PUPIALES | |
| All prom_pruebas | 48 | 50 | 47 | 47 | 47 | 50 | 50 | 54 | 48 | 50 | 50 | 51 | 51 | |

Slicer:



Fuente: Esta investigación

En la figura 24 se observa que los estudiantes con mejor promedio en la prueba de sociales ciudadanas son del municipio de Gualmatán.

Figura 25. Promedios de desempeño de la Prueba de ciencias naturales discriminado por municipios.

| | cole_mcpio_ubicacion | | | | | | | | | | | | | | |
|------------------|----------------------|-----------|---------|---------|-----------|--------|-------|-----------|-----------|------|---------|--------|---------|----------|----|
| prom_prueba | ALDANA | CONTADERO | CORDOBA | CUASPUD | CARLOSAMA | CUMBAL | FUNES | GUACHUCAL | GUAIMATAN | ILES | IPIALES | POTOSI | PUERRES | PUPIALES | |
| All prom_pruebas | 51 | 52 | 48 | | | 49 | 49 | 50 | 53 | 54 | 49 | 51 | 52 | 52 | 51 |

Slicer:



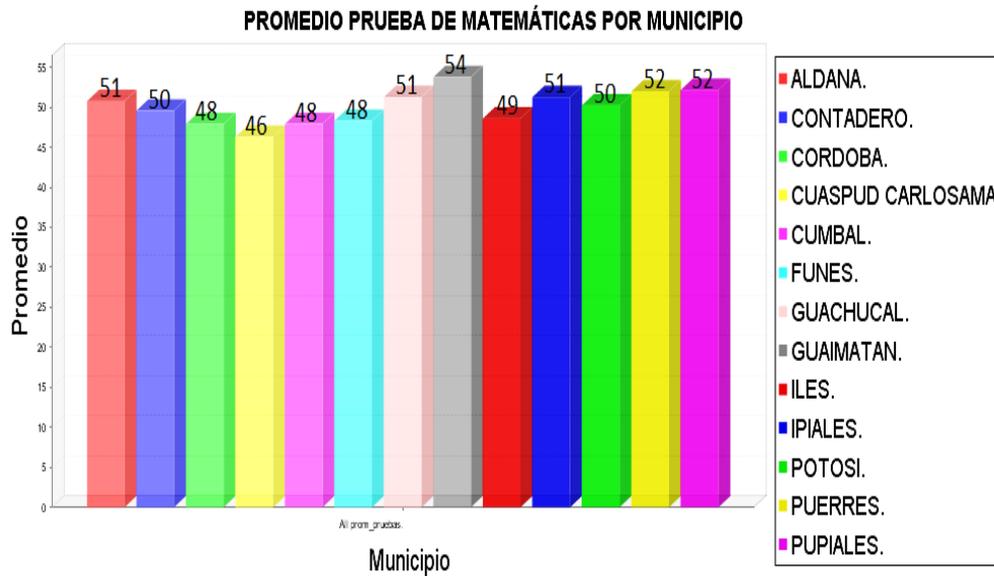
Fuente: Esta investigación

En la figura 25 se observa que los estudiantes con el menor promedio en la prueba de ciencias naturales son los del municipio de Córdoba.

Figura 26. Promedios de desempeño de la Prueba de matemáticas discriminado por municipios.

| | cole_mcpio_ubicacion | | | | | | | | | | | | | |
|------------------|----------------------|-----------|---------|-------------------|--------|-------|-----------|-----------|------|---------|--------|---------|----------|--|
| prom_prueba | ALDANA | CONTADERO | CORDOBA | CUASPUD CARLOSAMA | CUMBAL | FUNES | GUACHUCAL | GUAIMATAN | ILES | IPIALES | POTOSI | PUERRES | PUPIALES | |
| All prom_pruebas | 51 | 50 | 48 | 46 | 48 | 48 | 51 | 54 | 49 | 51 | 50 | 52 | 52 | |

Slicer:



Fuente: Esta investigación

A pesar de que todos los municipios de la subregión de Obando tienen un bajo promedio en todas las pruebas, cabe resaltar que los estudiantes de Gualmatán se destacan con unos mayores promedios que los demás municipios.

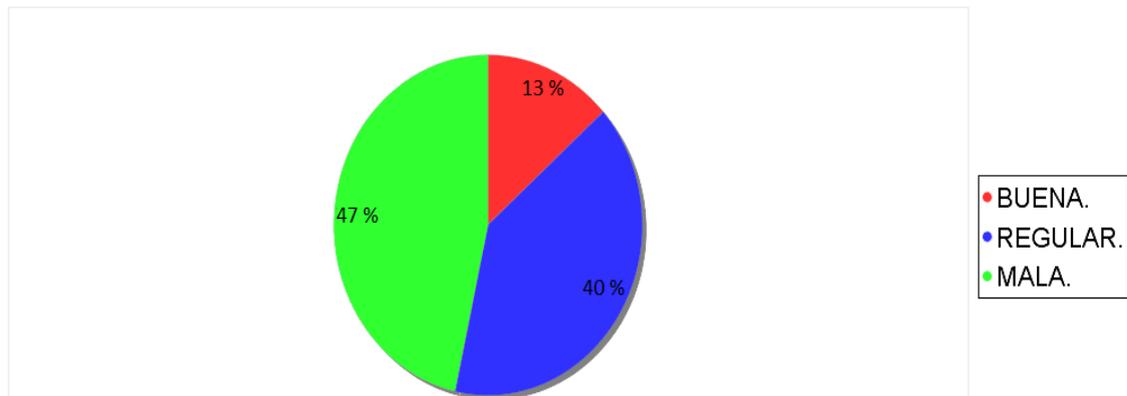
- ¿Cuántos estudiantes presentaron la Prueba Saber 11° por condición TIC?

Figura 27. Número de estudiantes que presentaron la prueba por condición TIC.

| | desemp_soc_ciu |
|-------------------|---------------------|
| eco_condicion_tic | All desemp_soc_cius |
| BUENA | 2430 |
| REGULAR | 7258 |
| MALA | 8431 |

Slicer:

No. ESTUDIANTES QUE PRESENTARON LA PRUEBA DE SOCIALES CIUDADANAS POR CONDICIÓN TIC



Fuente: Esta investigación

En la figura 27 se observa que del 2012 al 2016 los estudiantes que presentaron la Prueba, la mayor parte de ellos tiene una mala condición TIC equivalentes al 47 %. Y que tan solo 13 % tiene buenas condiciones.

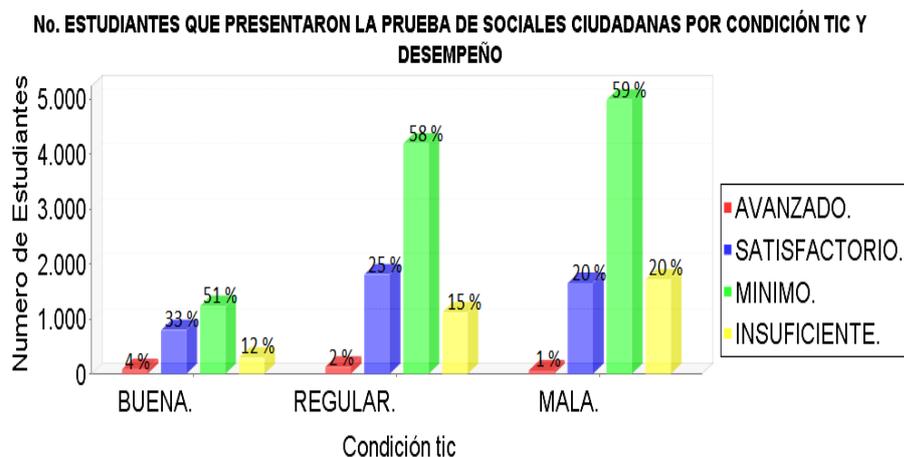
De tal forma que el alto porcentaje de la mala condición TIC en la subregión de Obando es preocupante, ya que las TIC son una herramienta muy importante para la educación.

- ¿Cuántos estudiantes presentaron la Prueba de sociales y ciudadanas discriminado por condición TIC y desempeño de la prueba?

Figura 28. Número de estudiantes que presentaron la prueba de sociales y ciudadanas por condición TIC y desempeño de la prueba.

| eco_condicion_tic | desemp_soc_ciu | | | |
|-------------------|----------------|---------------|--------|--------------|
| | AVANZADO | SATISFACTORIO | MINIMO | INSUFICIENTE |
| BUENA | 90 | 801 | 1240 | 299 |
| REGULAR | 132 | 1813 | 4193 | 1120 |
| MALA | 65 | 1656 | 4989 | 1721 |

Slicer:

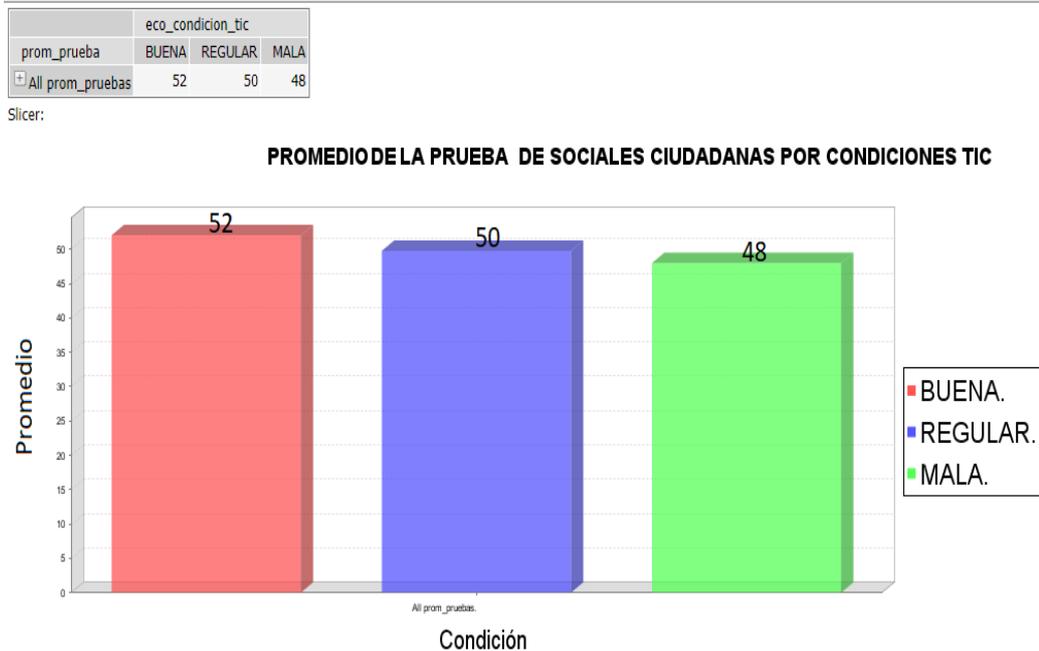


Fuente: Esta investigación

En la gráfica 28 se observa que con una buena condición TIC los estudiantes alcanzan un mejor rendimiento académico.

- ¿Cuáles son los promedios de la Prueba de sociales ciudadanas, matemáticas y ciencias naturales discriminado por condición TIC y promedio?

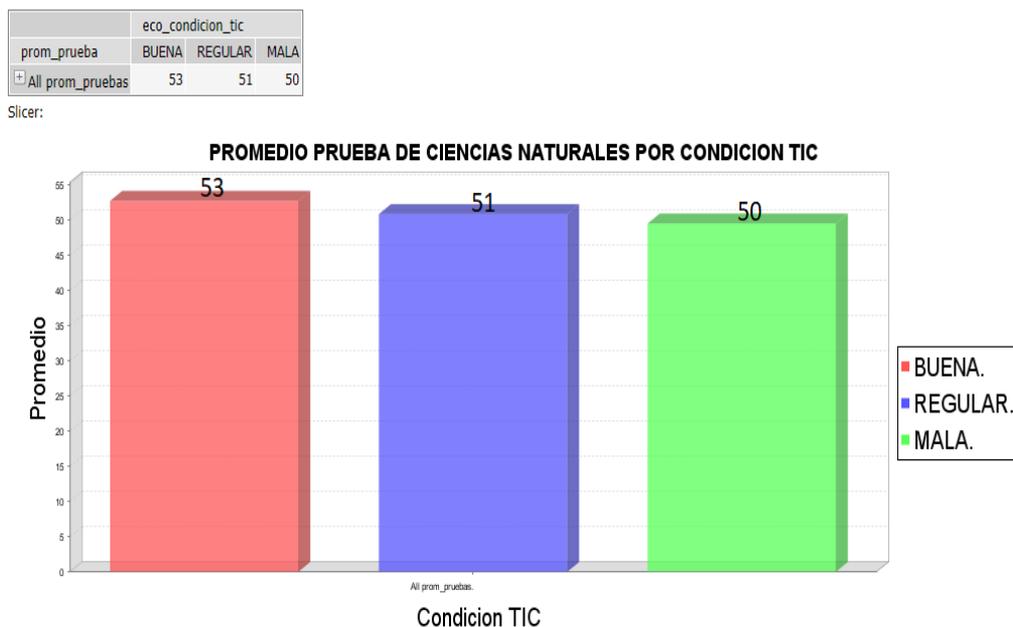
Figura 29. Promedio de estudiantes que presentaron la prueba de sociales ciudadanas por condición TIC.



Fuente: Esta investigación

En la figura 29 se observa que los estudiantes con mejor promedio en la prueba de sociales ciudadanas tienen una buena condición TIC.

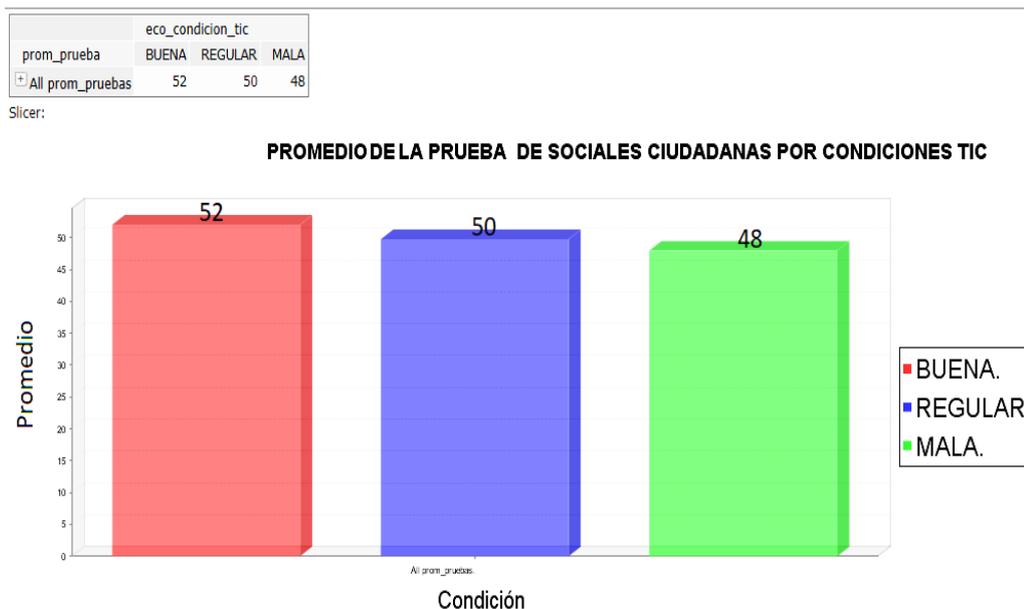
Figura 30. Promedio de estudiantes que presentaron la prueba de ciencias naturales por condición TIC y promedio.



Fuente: Esta investigación

En la figura 30 se observa que los estudiantes con el menor promedio en la prueba de ciencias naturales tienen una mala condición TIC.

Figura 31. Promedio de estudiantes que presentaron la prueba de matemáticas por condición TIC y promedio.



Fuente: Esta investigación

A pesar de que todas las condiciones TIC tienen un bajo promedio en las pruebas, cabe resaltar que los estudiantes con una buena condición TIC tienen mejor promedio que las demás.

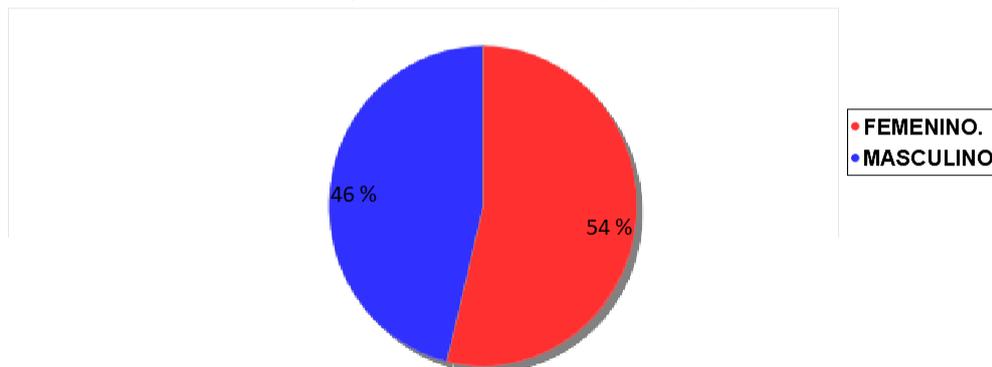
- ¿Cuántos estudiantes presentaron la prueba por género?

Figura 32. Número de estudiantes que presentaron la prueba por género.

| estu_genero | ano |
|-------------|------|
| FEMENINO | 9744 |
| MASCULINO | 8375 |

Slicer:

No. ESTUDIANTES QUE PRESENTARON LA PRUEBA POR GÉNERO



Fuente: Esta investigación

En la figura 32 se observa que en la subregión de Obando la mayor parte de los estudiantes que presentaron la prueba del 2012 al 2016 son mujeres equivalentes al 54 % y que el 46 % son hombres.

- ¿Cuántos estudiantes presentaron la prueba por año y género?

Figura 33. Número de estudiantes que presentaron la prueba por año y género.

| ano | estu_genero | |
|------|-------------|-----------|
| | FEMENINO | MASCULINO |
| 2012 | 2045 | 1871 |
| 2013 | 2096 | 1782 |
| 2014 | 2072 | 1748 |
| 2015 | 1795 | 1478 |
| 2016 | 1736 | 1496 |

Slicer:



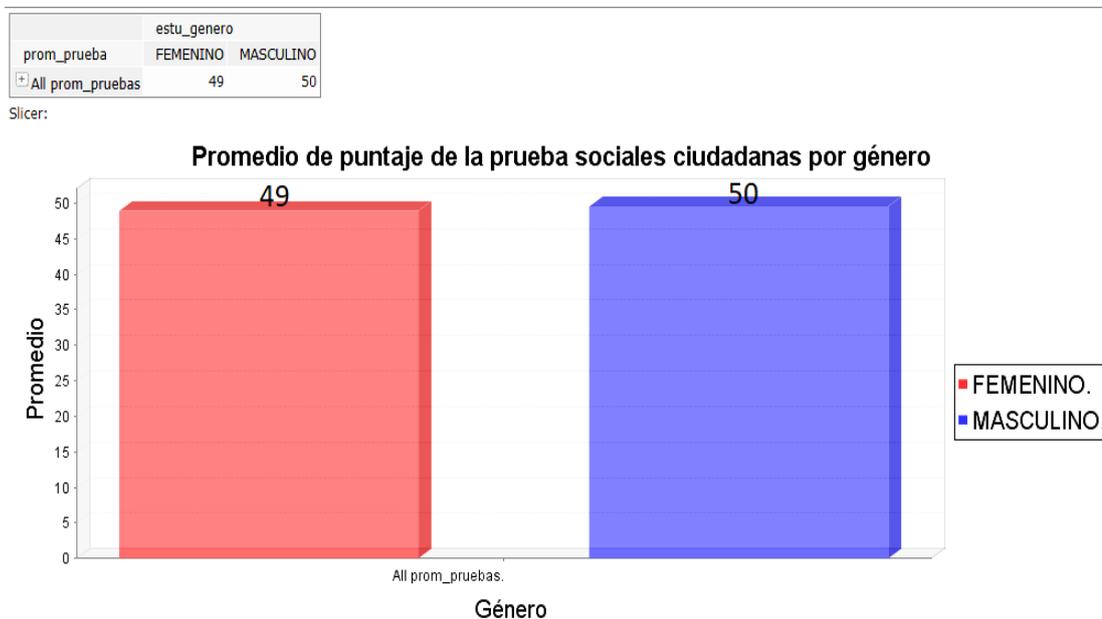
Fuente: Esta investigación

En la figura 33 se observa que del 2012 al 2016 existe mayor participación en estas pruebas por parte del género femenino en la subregión de Obando, pero también se puede afirmar que la población femenina como masculina disminuye poco a poco.

También se concluye que cada año el género masculino hace menos presencia en estas pruebas con respecto al género femenino.

- ¿Promedio de la Prueba de sociales y ciudadanas, ciencias naturales, matemáticas en el período del 2012 al 2016 discriminado por género?

Figura 34. Número de estudiantes que presentaron la prueba de sociales ciudadanas por género.



Fuente: Esta investigación

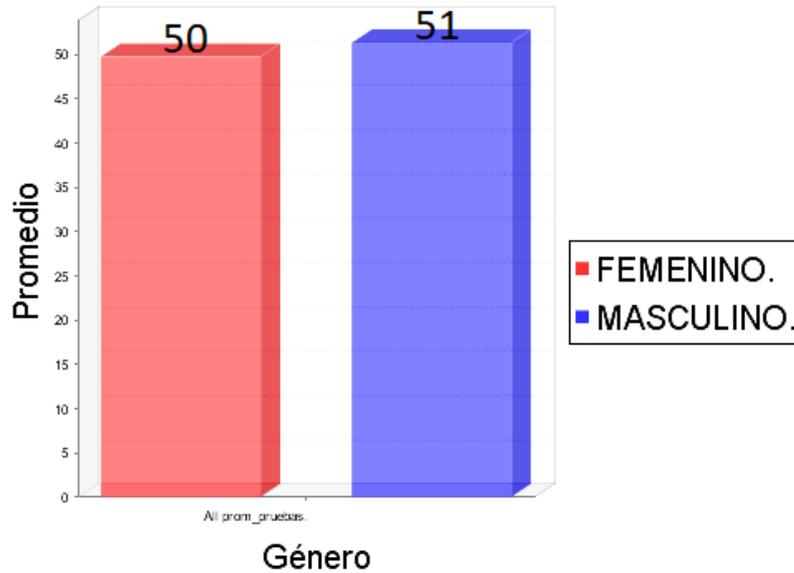
En la figura 34 se observa que en la prueba de sociales y ciudadanas, el género masculino tiene mejor promedio que el género femenino.

Figura 35. Número de estudiantes que presentaron la prueba de ciencias naturales por género.

| | estu_genero | |
|------------------|-------------|-----------|
| prom_prueba | FEMENINO | MASCULINO |
| All prom_pruebas | 50 | 51 |

Slicer:

PROMEDIO PRUEBA DE CIENCIAS NATURALES POR GÉNERO



Fuente: Esta investigación

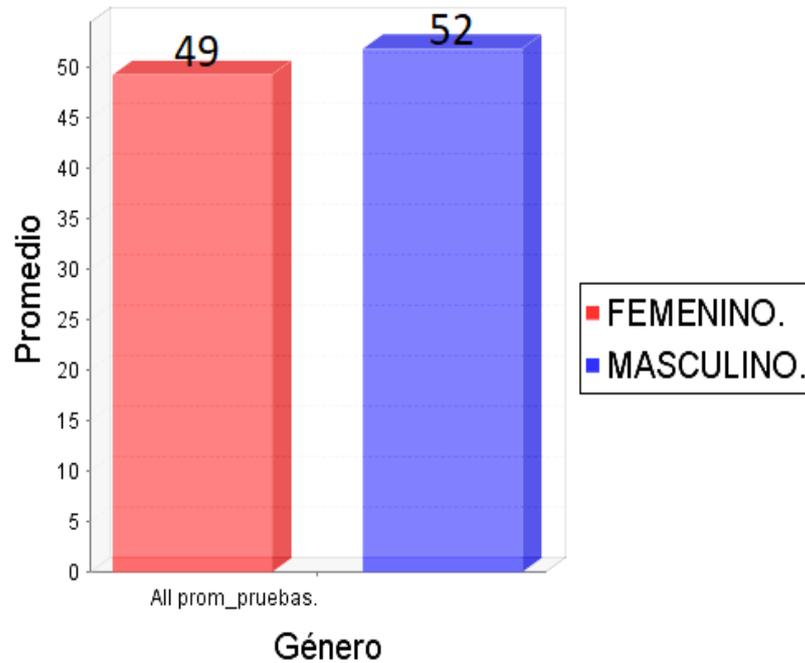
En la figura 35 se observa que en la prueba de ciencias naturales el género femenino tiene menor promedio que el género masculino.

Figura 36. Número de estudiantes que presentaron la prueba de matemáticas por género.

| prom_prueba | estu_genero | |
|--------------------|-------------|-----------|
| | FEMENINO | MASCULINO |
| + All prom_pruebas | 49 | 52 |

Slicer:

PROMEDIO PRUEBA DE MATEMÁTICAS POR GÉNERO



Fuente: Esta investigación

En la figura 36 se observa que el género masculino tiene un mejor promedio que el género femenino.

Se concluye que a pesar de que el género masculino haga menos presencia cada año, este presenta mejor promedio en todas las pruebas.

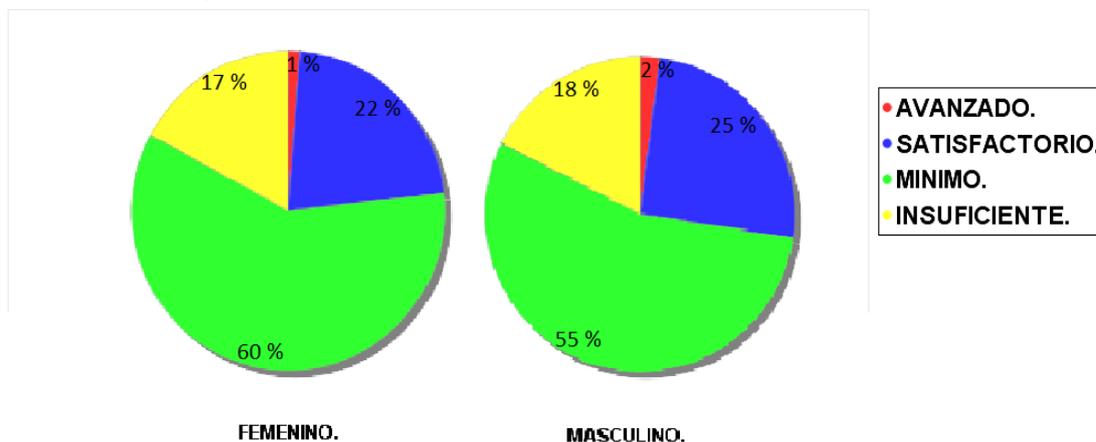
- ¿Cuántos estudiantes presentaron la Prueba de sociales y ciudadanas en el período del 2012 al 2016 discriminado por desempeño y género?

Figura 37. Número de estudiantes que presentaron la prueba de sociales ciudadanas por desempeño y género.

| desemp_soc_ciu | estu_genero | |
|----------------|-------------|-----------|
| | FEMENINO | MASCULINO |
| AVANZADO | 119 | 168 |
| SATISFACTORIO | 2144 | 2126 |
| MINIMO | 5811 | 4611 |
| INSUFICIENTE | 1670 | 1470 |

Slicer:

No. ESTUDIANTES QUE PRESENTARON LA PRUEBA DE SOCIALES CIUDADANAS POR DESEMPEÑO Y GÉNERO



Fuente: Esta investigación

En la figura 37 se observa que en la subregión de Obando tanto como para el género masculino y femenino el desempeño mínimo en la prueba de sociales y ciudadanas es demasiado alto.

La suma del desempeño mínimo e insuficiente en el género femenino equivalente al 76 %, el cual es muy preocupante ya que es muy alto el nivel de bajo rendimiento con respecto a esta prueba.

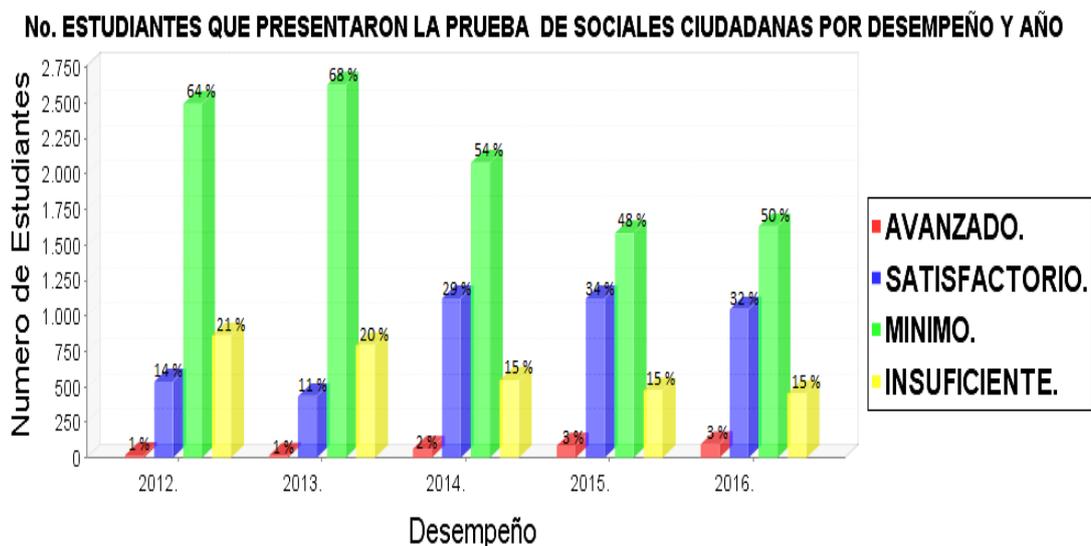
El género masculino a pesar de que tiene un mejor rendimiento en esta prueba que el género femenino no deja de ser preocupante, ya que la suma del desempeño mínimo e insuficiente en el género masculino equivalente al 72 %.

- ¿Cuántos estudiantes presentaron la Prueba de sociales y ciudadanas por desempeño y año?

Figura 38. Número de estudiantes que presentaron la prueba de sociales ciudadanas discriminado por desempeño y año.

| desemp_soc_ciu | | | | |
|----------------|----------|---------------|--------|--------------|
| ano | AVANZADO | SATISFACTORIO | MINIMO | INSUFICIENTE |
| 2012 | 20 | 538 | 2495 | 863 |
| 2013 | 17 | 435 | 2631 | 795 |
| 2014 | 65 | 1125 | 2081 | 549 |
| 2015 | 88 | 1125 | 1583 | 477 |
| 2016 | 97 | 1047 | 1632 | 456 |

Slicer:



Fuente: Esta investigación

En la figura 38 se observa que el nivel de desempeño mínimo es el más relevante en todos los años, aunque cabe resaltar que del 2014 al 2016 este nivel tiene una disminución favorable para la educación.

- ¿Cuántos estudiantes presentaron la Prueba de sociales y ciudadanas por desempeño?

Figura 39. Número de estudiantes que presentaron la prueba de sociales ciudadanas por desempeño.

| | | desemp_soc_ciu | | | |
|------------------|--|----------------|---------------|--------|--------------|
| estu_genero | | AVANZADO | SATISFACTORIO | MINIMO | INSUFICIENTE |
| All estu_generos | | 287 | 4270 | 10422 | 3140 |

Slicer:



En la figura 39 se observa el problema de bajo rendimiento en la educación en la subregión de Obando. Del 2012 al 2016 tan solo el 2 % de los estudiantes alcanzaron el desempeño avanzado.

¿Cuántos estudiantes presentaron la Prueba Saber 11° en el período del 2012 al 2016 discriminado por estrato de vivienda?

Figura 40. Número de estudiantes que representaron la prueba de estado.



Fuente: Esta investigación

En la figura 40 se observa que del 2012 al 2016 los estudiantes que presentaron la Prueba en la Subregión de Obando, la mayor parte de ellos fueron de estrato uno, equivalentes al 79 %, seguido por los estudiantes de estrato dos, equivalentes al 14 %. Tan solo el 6 % fueron de estratos superiores al estrato dos.

3. CONCLUSIONES

Con la implementación y desarrollo de este proyecto se logró diseñar y construir un mercado de datos con información limpia y confiable a través de las bases de datos y documentos de apoyo que brinda el Icfes a los investigadores.

La metodología planteada por Kimball fue la más acorde para el desarrollo de esta investigación, ya que esta permite implementar mercados de datos con pocos recursos en áreas específicas y de poco irlos integrando en una bodega de datos.

Como resultado de este trabajo de investigación, la Subregión de Obando contó con un mercado de datos que almacena la información histórica de los resultados de los estudiantes de los municipios de la subregión de Obando del departamento de Nariño, que presentaron las pruebas Saber 11° entre los años 2012 y 2016, el cual generó información de calidad para soportar la toma de decisiones acertadas en pro del mejoramiento de la calidad de la educación básica y media de esta subregión.

La etapa de la recolección de información fue clave en el desarrollo de la investigación en la cual se utilizó herramientas de software libre como lo es PostgreSQL, permitiendo esta, integrar en una base de datos inicial la información de las pruebas Saber 11°.

La construcción de la base de datos inicial permitió el análisis de la información, creando así el repositorio temporal `saber11_2012_2016` con atributos socioeconómicos, académicos e institucionales de cada estudiante de las instituciones educativas de la Subregión de Obando.

La construcción e implementación del modelo multidimensional en estrella, permitió mediante la herramienta de Inteligencia de Negocios, libre Pentaho; analizar y evaluar informes y reportes generados.

Dentro de este proyecto se aplicó los conocimientos adquiridos durante toda la formación académica junto con el proceso investigativo desarrollado en el grupo GRIAS. Es significativo también reconocer la importancia que tiene para los estudiantes el hecho de pertenecer al grupo de investigación debido a la experiencia adquirida y la constante asesoría por parte de los docentes e investigadores del grupo.

4. RECOMENDACIONES

Se recomienda a los municipios de la Subregión de Obando actualizar el mercado de datos con los subsiguientes años.

Para obtener resultados confiables es fundamental realizar un análisis serio de la calidad inicial de los datos, con el fin de establecer los procedimientos necesarios para realizar la limpieza y estandarización de éstos.

En la etapa de extracción, transformación y carga de los datos se recomienda aplicar los recursos y tiempo necesarios para garantizar el éxito del proyecto.

Se recomienda documentar cada una de las fases del proyecto y generar los metadatos correspondientes con el fin de poder replicar el proceso en trabajos futuros de la misma naturaleza.

Se recomienda que el gobierno nacional utilice tecnología de inteligencia de negocios para establecer políticas serias de mejoramiento de los programas académicos.

Se recomienda a cada una de las instituciones educativas de la subregión de Obando hacer uso de este mercado de datos el cual constituye una poderosa herramienta de inteligencia de negocios que permite la construcción de reportes a partir del análisis multidimensional, con grandes posibilidades de variar el foco de visualización, generando nuevo conocimiento que permitió tomar decisiones óptimas.

Para que los municipios de la Subregión de Obando puedan utilizar este mercado de datos fácilmente, como trabajo futuro se recomienda desarrollar una aplicación web, que le facilite los diferentes procesos y la actualización del mercado de datos.

A trabajos futuros en relación con este mercado de datos se recomienda la aplicación de minería de datos para descubrir patrones que mejoren la calidad de educación.

BIBLIOGRAFÍA

Fernández, H. Como Interpretar la Evaluación Pruebas Saber. Ministerio de Educación Nacional. Bogotá, Colombia.2005.

Posada, J. y Mendoza, F. Determinantes del logro académico de los estudiantes de grado 11 en el periodo 2008 2010. Una perspectiva de género y región. Estudios sobre calidad de la educación en Colombia, ICFES, Ministerio de Educación. Bogotá, Colombia. 2014

ICFES. Sistema Nacional de Evaluación Estandarizada de la Educación Alineación del examen SABER 11°.Lineamientos generales 2014 - 2. ISBN: 978-958-11-0630-1. Bogotá, Colombia. 2014.

Gobernación de Nariño, Ex-Provincia de Obando, 26-11-2012. [En línea].Avalible: <http://xn--nario-rta.gov.co/2012-2015/index.php/ex-providencia-de-obando>

ICFES, Consulta de resultados, pruebas saber 3° 5° 9° 11° [En línea]. Avaliable: <http://www2.icfesinteractivo.gov.co/ReportesSaber359/consultaReporteEntidadTerritorial.aspx>. 2016.

ISIS, Diseño y construcción de un data mart para la mantención de indicadores de sostenibilidad de la industria del salmón. [En line]. Available: https://www.u-cursos.cl/ingenieria/2009/1/IN830/1/material_docente/bajar?

INEGI. Diseño de un modelo multidimensional de data mart del área de capacitación para el INEGI. Universidad Autónoma de Aguascalientes. 2008.[En línea]. Available: <http://bdigital.dgse.uaa.mx:8080/xmlui/bitstream/handle/123456789/526/299008.pdf?sequence=1>.

ECCI. DATAMART para apoyo a las estrategias de permanencia estudiantil en la Escuela Colombiana de carreras Industriales. 2013. [En línea]. Available:http://www.colombiaaprende.edu.co/html/micrositios/1752/articles-344993_recurso_3.pdf.

Arenas, M. y Gomez, A. Inteligencia de Negocios aplicada a los procesos de autoevaluación de la Universidad de Manizales. Universidad de Manizales. 2017. [En línea].Available :http://ridum.umanizales.edu.co:8080/xmlui/bitstream/handle/6789/3046/Gomez_Montes_Ana_Maria_2017.pdf?

Barreto, J. y Bocanegra, D. Diseño de un Data Mart de información académica para la Universidad de Nariño. Departamento de Sistemas, Facultad de Ingeniería, 2009. Proceso Software. Madrid: Ramon Areces.

T. Erl, W. Khattak, y P. Buhler, Big Data Fundamentals. Concepts, Drivers & Techniques. Prentice Hall.

Imhoff and Galemno, Mastering Data Warehouse Design: Relational and Dimensional Techniques, Wiley Publishing, 2003.

R. Kimball y A. Ross, The data warehouse toolkit, Third. Indianapolis: John Wiley & Sons, 2013.

C. Ballard, A. Gupta, V. Krishnan, N. Pessoa, y O. Stephan, Data Mart Consolidation: Getting Control of Your Enterprise Information, First. IBM, 2005.

E. Malinowski y E. Zimányi, Advanced data warehouse design. From Conventional to Spatial and Temporal Applications. Berlin: Springer-Verlag, 2008.

Inmon, Building the Data Warehouse, John Wiley & Sons, 2002.

Rivadeneira, G, La metodología de Kimball para el diseño de almacenes de datos (Data warehouses), en Cuadernos de la Facultad de Ingeniería e Informática Número 5, Universidad Católica de Salta, 2010.

M. Reyes y P. Rosales, “Desarrollo de un datamart de información académica de estudiantes de la Escuela de Ciencias y Sistemas de la Facultad de Ingeniería de la USAC”, Trabajo de grado para optar al título de Ingeniero en Ciencias y Sistemas, Guatemala, 2007.

A. Rojas, “Implementación de un data mart como solución de inteligencia de negocios, bajo la metodología de Ralph Kimball para optimizar la toma de decisiones en el Departamento de Finanzas de la Contraloría General de la República”, Tesis para optar al título profesional de Ingeniero de Computación y Sistemas, Chiclayo, 2014

Congreso de Colombia. (13 de julio del 2009). artículo 12. Ministerio de Educación Nacional. [LEY 1324 DE 2009]

Congreso de Colombia. (13 de julio del 2009). artículo 7. Ministerio de Educación Nacional. [LEY 1324 DE 2009]

Instituto Colombiano para la Evaluación de la Educación ICFES. (2016). El antes y ahora del examen. [En línea] .Available: <http://www.icfes.gov.co/instituciones-educativas-y-secretarias/saber-11/el-antes-y-ahora-del-examen>

Instituto Colombiano para la Evaluación de la Educación ICFES .tipos de resultados generados a partir de las evaluaciones. [En línea]. Available:http://www2.icfes.gov.co/index.php?option=com_k2&view=item&layout=item&id=541&Itemid=328

PostgreSQL, «About,»[Enlínea]. Available:<https://www.postgresql.org/about/>.

RodryFuyol's Blog. ¿Qué es Pentaho y para qué sirve? [En línea]. Available: <https://rodryfuyol.wordpress.com/queespentahoyparaquesirve/>

Factores asociados. ICFES. 2017. [En línea]. Available <http://www.icfes.gov.co/instituciones-educativas-y-secretarias/pruebas-saber-3-5-y-9/factores-asociados>

ANEXOS

1. SCRIPTS:

```
CREATE TABLE saber11_2012_2016_limpio AS SELECT * FROM  
saber11_2012_2016;
```

2. SCRIPTS:

```
UPDATE saber11_2012_2016_limpio  
SET estu_reside_mcpio='BOGOTA D.C'  
WHERE estu_reside_mcpio=' BOGOTÁ D.C.';
```

```
UPDATE saber11_2012_2016_limpio  
SET estu_reside_mcpio='BOJACA'  
WHERE estu_reside_mcpio=' BOJACÁ';
```

```
UPDATE saber11_2012_2016_limpio  
SET estu_reside_mcpio='BRICENO'  
WHERE estu_reside_mcpio='BRICEÑO';
```

```
UPDATE saber11_2012_2016_limpio  
SET estu_reside_mcpio=' CANASGORDAS'  
WHERE estu_reside_mcpio='CAÑASGORDAS';
```

```
UPDATE saber11_2012_2016_limpio  
SET estu_reside_mcpio='COVENAS '  
WHERE estu_reside_mcpio='COVEÑAS';
```

```
UPDATE saber11_2012_2016_limpio  
SET estu_reside_mcpio='EL PENOL'  
WHERE estu_reside_mcpio='EL PEÑOL';
```

```
UPDATE saber11_2012_2016_limpio  
SET estu_reside_mcpio='EL PENON'  
WHERE estu_reside_mcpio='EL PEÑON';
```

```
UPDATE saber11_2012_2016_limpio  
SET estu_reside_mcpio='EL PINON'  
WHERE estu_reside_mcpio='EL PIÑON';
```

```
UPDATE saber11_2012_2016_limpio  
SET estu_reside_mcpio='GUACHENE'  
WHERE estu_reside_mcpio='GUACHENÉ';
```

```
UPDATE saber11_2012_2016_limpio
SET estu_reside_mcpio='LA MONTANITA'
WHERE estu_reside_mcpio='LA MONTAÑITA';
```

```
UPDATE saber11_2012_2016_limpio
SET estu_reside_mcpio='LA PENA'
WHERE estu_reside_mcpio='LA PEÑA';
```

```
UPDATE saber11_2012_2016_limpio
SET estu_reside_mcpio='MONITOS'
WHERE estu_reside_mcpio='MOÑITOS';
```

```
UPDATE saber11_2012_2016_limpio
SET estu_reside_mcpio='NARINO'
WHERE estu_reside_mcpio='NARIÑO';
```

```
UPDATE saber11_2012_2016_limpio
SET estu_reside_mcpio='OCANA'
WHERE estu_reside_mcpio='OCAÑA';
```

```
UPDATE saber11_2012_2016_limpio
SET estu_reside_mcpio='PENOL'
WHERE estu_reside_mcpio='PEÑOL';
```

```
UPDATE saber11_2012_2016_limpio
SET estu_reside_mcpio='PIJINO DEL CARMEN'
WHERE estu_reside_mcpio='PIJIÑO DEL CARMEN';
```

```
UPDATE saber11_2012_2016_limpio
SET estu_reside_mcpio='PUERTO NARINO'
WHERE estu_reside_mcpio='PUERTO NARIÑO';
```

```
UPDATE saber11_2012_2016_limpio
SET estu_reside_mcpio='SALDANA'
WHERE estu_reside_mcpio='SALDAÑA';
```

```
UPDATE saber11_2012_2016_limpio
SET estu_reside_mcpio='SAN JOSE DE LA MONTANA'
WHERE estu_reside_mcpio='SAN JOSE DE LA MONTAÑA';
```

```
UPDATE saber11_2012_2016_limpio
SET estu_reside_mcpio='SAN JOSE DE URE'
```

```
WHERE estu_reside_mcpio='SAN JOSÉ DE URÉ';
```

```
UPDATE saber11_2012_2016_limpio  
SET estu_reside_mcpio='SANTAFE DE ANTIOQUIA'  
WHERE estu_reside_mcpio='SANTAFÉ DE ANTIOQUIA';
```

```
UPDATE saber11_2012_2016_limpio  
SET estu_reside_mcpio='TUCHIN'  
WHERE estu_reside_mcpio='TUCHÍN';
```

3. SCRIPTS:

```
UPDATE saber11_2012_2016_limpio  
SET estu_reside_depto='ATLANTICO'  
WHERE estu_reside_depto='ATLÁNTICO';
```

```
UPDATE saber11_2012_2016_limpio  
SET estu_reside_depto='BOGOTA'  
WHERE estu_reside_depto='BOGOTÁ';
```

```
UPDATE saber11_2012_2016_limpio  
SET estu_reside_depto='BOLIVAR'  
WHERE estu_reside_depto='BOLÍVAR';
```

```
UPDATE saber11_2012_2016_limpio  
SET estu_reside_depto='BOYACA'  
WHERE estu_reside_depto='BOYACÁ';
```

```
UPDATE saber11_2012_2016_limpio  
SET estu_reside_depto='CAQUETA'  
WHERE estu_reside_depto='CAQUETÁ';
```

```
UPDATE saber11_2012_2016_limpio  
SET estu_reside_depto='CHOCO'  
WHERE estu_reside_depto='CHOCÓ';
```

```
UPDATE saber11_2012_2016_limpio  
SET estu_reside_depto='CORDOBA'  
WHERE estu_reside_depto='CÓRDOBA';
```

```
UPDATE saber11_2012_2016_limpio  
SET estu_reside_depto='GUAINIA'  
WHERE estu_reside_depto='GUAINÍA';
```

```
UPDATE saber11_2012_2016_limpio
SET estu_reside_depto='NARINO'
WHERE estu_reside_depto='NARIÑO';
```

```
UPDATE saber11_2012_2016_limpio
SET estu_reside_depto='QUINDIO'
WHERE estu_reside_depto='QUINDÍO';
```

```
UPDATE saber11_2012_2016_limpio
SET estu_reside_depto='SAN ANDRES'
WHERE estu_reside_depto='SAN ANDRÉS';
```

```
UPDATE saber11_2012_2016_limpio
SET estu_reside_depto='VAUPES'
WHERE estu_reside_depto='VAUPÉS';
```

4. SCRIPTS:

```
UPDATE saber11_2012_2016_limpio
SET COLE_JORNADA = 'MANANA'
WHERE COLE_JORNADA = 'MAÑANA';
```

```
UPDATE saber11_2012_2016_limpio
SET COLE_JORNADA = 'SABATINA_DOMINICAL'
WHERE COLE_JORNADA = 'SABATINA - DOMINICAL';
```

5. SCRIPTS:

```
UPDATE saber11_2012_2016_limpio
SET cole_jornada='UNICA'
WHERE cole_jornada='U';
```

6. SCRIPTS:

```
UPDATE saber11_2012_2016_limpio
SET cole_jornada='SABATINA DOMINICAL'
WHERE cole_jornada='S';
```

7. SCRIPTS:

```
UPDATE saber11_2012_2016_limpio
SET cole_jornada='COMPLETA U ORDINARIA'
WHERE cole_jornada='C';
```

8. SCRIPTS:

```
UPDATE saber11_2012_2016_limpio
SET cole_jornada='MANANA'
WHERE cole_jornada='M';
```

9. SCRIPTS:

```
UPDATE saber11_2012_2016_limpio
SET cole_jornada='TARDE'
WHERE cole_jornada='T';
```

10. SCRIPTS:

```
UPDATE saber11_2012_2016_limpio
SET cole_jornada='NOCHE'
WHERE cole_jornada='N';
```

11. SCRIPTS:

```
UPDATE saber11_2012_2016_limpio
SET COLE_CHARACTER = 'ACADEMICO'
WHERE COLE_CHARACTER ='ACADÉMICO';
```

```
UPDATE saber11_2012_2016_limpio
SET COLE_CHARACTER = 'TECNICO'
WHERE COLE_CHARACTER ='TÉCNICO';
```

```
UPDATE saber11_2012_2016_limpio
SET COLE_CHARACTER = 'ACADEMICO Y TECNICO'
WHERE COLE_CHARACTER ='TÉCNICO/ACADÉMICO';
```

12. SCRIPTS:

```
UPDATE saber11_2012_2016_limpio
SET estu_area_reside='U'
WHERE estu_area_reside='1';
```

13. SCRIPTS:

```
UPDATE saber11_2012_2016_limpio
SET estu_area_reside='R'
WHERE estu_area_reside='2';
```

14. SCRIPTS:

```
UPDATE saber11_2012_2016_limpio
SET cole_genero ='FEMENINO'
WHERE cole_genero ='F';
```

15. SCRIPTS:
UPDATE saber11_2012_2016_limpio
SET cole_genero ='MASCULINO'
WHERE cole_genero ='M';

16. SCRIPTS:
UPDATE saber11_2012_2016_limpio
SET cole_genero ='MIXTO'
WHERE cole_genero ='X';

17. SCRIPTS:
UPDATE saber11_2012_2016_limpio
SET cole_naturaleza ='OFICIAL'
WHERE cole_naturaleza='O';

18. SCRIPTS:
UPDATE saber11_2012_2016_limpio
SET cole_naturaleza ='NO OFICIAL'
WHERE cole_naturaleza='N';

19. SCRIPTS:
UPDATE saber11_2012_2016_limpio
SET fami_dvd = '0'
WHERE fami_dvd= 'N';

20. SCRIPTS:
UPDATE saber11_2012_2016_limpio
SET fami_dvd = '1'
WHERE fami_dvd= 'S';

21. SCRIPTS:
UPDATE saber11_2012_2016_limpio
SET FAMI_TELEFONO_FIJO = '0'
WHERE FAMI_TELEFONO_FIJO= 'N';

22. SCRIPTS:
UPDATE saber11_2012_2016_limpio
SET FAMI_TELEFONO_FIJO = '1'
WHERE FAMI_TELEFONO_FIJO= 'S';

23. SCRIPTS:
UPDATE saber11_2012_2016_limpio

```
SET FAMI_TELEVISOR = '0'  
WHERE FAMI_TELEVISOR = 'N';
```

24. SCRIPTS:

```
UPDATE saber11_2012_2016_limpio  
SET FAMI_TELEVISOR = '1'  
WHERE FAMI_TELEVISOR = 'S';
```

25. SCRIPTS:

```
UPDATE saber11_2012_2016_limpio  
SET FAMI_INTERNET = '0'  
WHERE FAMI_INTERNET = 'N';
```

26. SCRIPTS:

```
UPDATE saber11_2012_2016_limpio  
SET FAMI_AUTOMOVIL = '1'  
WHERE FAMI_AUTOMOVIL = 'S';
```

27. SCRIPTS:

```
UPDATE saber11_2012_2016_limpio  
SET fami_nivel_sisben = 'NIVEL 1'  
Where fami_nivel_sisben = '1';
```

```
UPDATE saber11_2012_2016_limpio  
SET fami_nivel_sisben = 'NIVEL 2'  
Where fami_nivel_sisben = '2';
```

```
UPDATE saber11_2012_2016_limpio  
SET fami_nivel_sisben = 'NIVEL 3'  
Where fami_nivel_sisben = '3';
```

```
UPDATE saber11_2012_2016_limpio  
SET fami_nivel_sisben = 'ESTA CLASIFICADA EN OTRO NIVEL DE SISBEN'  
Where fami_nivel_sisben = '4';
```

```
UPDATE saber11_2012_2016_limpio  
SET fami_nivel_sisben = 'NO ESTA CLASIFICADA POR EL SISBEN'  
Where fami_nivel_sisben = '5';
```

28. SCRIPTS:

```
UPDATE saber11_2012_2016_limpio  
SET fami_ingreso_fmiliar_mensual = 'MENOS DE 1 SM'
```

```
WHERE fami_ingreso_fmiliar_mensual = '1';
```

```
UPDATE saber11_2012_2016_limpio  
SET fami_ingreso_fmiliar_mensual= 'ENTRE 1 Y MENOS DE 2 SM'  
WHERE fami_ingreso_fmiliar_mensual = '2';
```

```
UPDATE saber11_2012_2016_limpio  
SET fami_ingreso_fmiliar_mensual= 'ENTRE 2 Y MENOS DE 3 SM'  
WHERE fami_ingreso_fmiliar_mensual = '3';
```

```
UPDATE saber11_2012_2016_limpio  
SET fami_ingreso_fmiliar_mensual= 'ENTRE 3 Y MENOS DE 5 SM'  
WHERE fami_ingreso_fmiliar_mensual = '4';
```

```
UPDATE saber11_2012_2016_limpio  
SET fami_ingreso_fmiliar_mensual= 'ENTRE 5 Y MENOS DE 7 SM'  
WHERE fami_ingreso_fmiliar_mensual = '5';
```

```
UPDATE saber11_2012_2016_limpio  
SET fami_ingreso_fmiliar_mensual= 'ENTRE 7 Y MENOS DE 10 SM'  
WHERE fami_ingreso_fmiliar_mensual = '6';
```

```
UPDATE saber11_2012_2016_limpio  
SET fami_ingreso_fmiliar_mensual= '10 O MAS SM'  
WHERE fami_ingreso_fmiliar_mensual = '7';
```

29. SCRIPTS:

```
CREATE TABLE saber11_2012_2016_transformado  
AS SELECT * FROM saber11_2012_2016_limpio;
```

30. SCRIPTS:

```
ALTER TABLE saber11_2012_2016_transformado  
ADD estu_edad_presentacion text;
```

31. SCRIPTS:

```
UPDATE saber11_2012_2016_transformado  
SET estu_edad_presentacion = estu_edad;
```

```
ALTER TABLE saber11_2012_2016_transformado  
ALTER COLUMN estu_edad_presentacion TYPE integer USING  
(trim(estu_edad_presentacion):integer);
```

32. SCRIPTS:

```
ALTER TABLE saber11_2012_2016_transformado
ADD estu_rango_edad text;
```

33. SCRIPTS:

```
UPDATE saber11_2012_2016_transformado
SET estu_rango_edad = '<17'
Where estu_edad_presentacion < 17;
```

```
UPDATE saber11_2012_2016_transformado
SET estu_rango_edad = '[17]'
Where estu_edad_presentacion = 17;
```

```
UPDATE saber11_2012_2016_transformado
SET estu_rango_edad = '[18-19]'
Where estu_edad_presentacion =18 or estu_edad_presentacion =19;
```

```
UPDATE saber11_2012_2016_transformado
SET estu_rango_edad = '[20-28]'
Where estu_edad_presentacion >= 20 and estu_edad_presentacion <=28;
```

```
UPDATE saber11_2012_2016_transformado
SET estu_rango_edad = '>28]'
Where estu_edad_presentacion > 28;
```

34. SCRIPTS:

```
ALTER TABLE saber11_2012_2016_transformado
ADD estu_discapacidad text;
```

35. SCRIPTS:

```
ALTER TABLE saber11_2012_2016_transformado
ADD fami_cod_max_nivel_educa_padres int;
```

36. SCRIPTS:

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN
fami_educa_padre TYPE integer USING (trim(fami_educa_padre)::integer);
```

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN
fami_educa_madre TYPE integer USING (trim(fami_educa_madre)::integer);
```

```
UPDATE saber11_2012_2016_transformado
```

```
SET fami_cod_max_nivel_educacion_padres = fami_educacion_padre;
```

```
UPDATE saber11_2012_2016_transformado  
SET fami_cod_max_nivel_educacion_padres = fami_educacion_madre  
Where fami_educacion_madre > fami_educacion_padre;
```

37. SCRIPTS:

```
ALTER TABLE saber11_2012_2016_transformado  
ADD fami_max_nivel_educacion_padres text;
```

38. SCRIPTS:

```
UPDATE saber11_2012_2016_transformado  
SET fami_max_nivel_educacion_padres = 'NINGUNO'  
Where fami_cod_max_nivel_educacion_padres = '1';
```

```
UPDATE saber11_2012_2016_transformado  
SET fami_max_nivel_educacion_padres = 'PRIMARIA INCOMPLETA'  
Where fami_cod_max_nivel_educacion_padres = '2';
```

```
UPDATE saber11_2012_2016_transformado  
SET fami_max_nivel_educacion_padres = 'PRIMARIA COMPLETA'  
Where fami_cod_max_nivel_educacion_padres = '3';
```

```
UPDATE saber11_2012_2016_transformado  
SET fami_max_nivel_educacion_padres = 'SECUNDARIA BACHILLERATO  
INCOMPLETO'  
Where fami_cod_max_nivel_educacion_padres = '4';
```

```
UPDATE saber11_2012_2016_transformado  
SET fami_max_nivel_educacion_padres = 'SECUNDARIA BACHILLERATO  
COMPLETO'  
Where fami_cod_max_nivel_educacion_padres = '5';
```

```
UPDATE saber11_2012_2016_transformado  
SET fami_max_nivel_educacion_padres = 'EDUCACION TECNICA O  
TECNOLOGICA INCOMPLETA'  
Where fami_cod_max_nivel_educacion_padres = '6';
```

```
UPDATE saber11_2012_2016_transformado  
SET fami_max_nivel_educacion_padres = 'EDUCACION TECNICA O  
TECNOLOGICA COMPLETA'  
Where fami_cod_max_nivel_educacion_padres = '7';
```

```
UPDATE saber11_2012_2016_transformado
SET fami_max_nivel_educacion_padres = 'EDUCACION PROFECIONAL
INCOMPLETA'
Where fami_cod_max_nivel_educacion_padres = '8';
```

```
UPDATE saber11_2012_2016_transformado
SET fami_max_nivel_educacion_padres = 'EDUCACION PROFECIONAL
COMPLETA'
Where fami_cod_max_nivel_educacion_padres = '9';
```

```
UPDATE saber11_2012_2016_transformado
SET fami_max_nivel_educacion_padres = 'POSTGRADO'
Where fami_cod_max_nivel_educacion_padres = '10';
```

```
UPDATE saber11_2012_2016_transformado
SET fami_max_nivel_educacion_padres = 'NO SABE'
Where fami_cod_max_nivel_educacion_padres = '11';
```

39. SCRIPTS:

```
ALTER TABLE saber11_2012_2016_transformado
ADD fami_ocup_padre text;
```

40. SCRIPTS:

```
ALTER TABLE saber11_2012_2016_transformado
ADD fami_ocup_madre text;
```

41. SCRIPTS:

```
UPDATE saber11_2012_2016_transformado
SET fami_ocup_madre = 'EMPRESARIO'
Where fami_ocupa_madre = '1';
```

```
UPDATE saber11_2012_2016_transformado
SET fami_ocup_madre = 'PEQUENO EMPRESARIO'
Where fami_ocupa_madre = '2';
```

```
UPDATE saber11_2012_2016_transformado
SET fami_ocup_madre = 'EMPLEADO CON CARGO COMO DIRECTOR O
GERENTE GENERAL'
Where fami_ocupa_madre = '3';
```

```
UPDATE saber11_2012_2016_transformado
```

```
SET fami_ocup_madre = 'EMPLEADO DE NIVEL DIRECTIVO'  
Where fami_ocupa_madre = '4';
```

```
UPDATE saber11_2012_2016_transformado  
SET fami_ocup_madre = 'EMPLEADO DE NIVEL TECNICO O PROFESIONAL'  
Where fami_ocupa_madre = '5';
```

```
UPDATE saber11_2012_2016_transformado  
SET fami_ocup_madre = 'EMPLEADO DE NIVEL AUXILIAR O  
ADMINISTRATIVO'  
Where fami_ocupa_madre = '6';
```

```
UPDATE saber11_2012_2016_transformado  
SET fami_ocup_madre = 'EMPLEADO OBRERO U OPERARIO'  
Where fami_ocupa_madre = '7';
```

```
UPDATE saber11_2012_2016_transformado  
SET fami_ocup_madre = 'PROFESIONAL INDEPENDIENTE'  
Where fami_ocupa_madre = '8';
```

```
UPDATE saber11_2012_2016_transformado  
SET fami_ocup_madre = 'TRABAJADOR POR CUENTA PROPIA'  
Where fami_ocupa_madre = '9';
```

```
UPDATE saber11_2012_2016_transformado  
SET fami_ocup_madre = 'HOGAR'  
Where fami_ocupa_madre = '10';
```

```
UPDATE saber11_2012_2016_transformado  
SET fami_ocup_madre = 'PENSIONADO'  
Where fami_ocupa_madre = '11';
```

```
UPDATE saber11_2012_2016_transformado  
SET fami_ocup_madre = 'OTRA ACTIVIDAD U OCUPACION'  
Where fami_ocupa_madre = '12';
```

```
UPDATE saber11_2012_2016_transformado  
SET fami_ocup_padre = 'EMPRESARIO'  
Where fami_ocupa_padre = '1';
```

```
UPDATE saber11_2012_2016_transformado  
SET fami_ocup_padre = 'PEQUENO EMPRESARIO'
```

Where fami_ocupa_padre = '2';

UPDATE saber11_2012_2016_transformado
SET fami_ocup_padre = 'EMPLEADO CON CARGO COMO DIRECTOR O GERENTE GENERAL'

Where fami_ocupa_padre = '3';

UPDATE saber11_2012_2016_transformado
SET fami_ocup_padre = 'EMPLEADO DE NIVEL DIRECTIVO'

Where fami_ocupa_padre = '4';

UPDATE saber11_2012_2016_transformado
SET fami_ocup_padre = 'EMPLEADO DE NIVEL TECNICO O PROFESIONAL'

Where fami_ocupa_padre = '5';

UPDATE saber11_2012_2016_transformado
SET fami_ocup_padre = 'EMPLEADO DE NIVEL AUXILIAR O ADMINISTRATIVO'

Where fami_ocupa_padre = '6';

UPDATE saber11_2012_2016_transformado
SET fami_ocup_padre = 'EMPLEADO OBRERO U OPERARIO'

Where fami_ocupa_padre = '7';

UPDATE saber11_2012_2016_transformado
SET fami_ocup_padre = 'PROFESIONAL INDEPENDIENTE'

Where fami_ocupa_padre = '8';

UPDATE saber11_2012_2016_transformado
SET fami_ocup_padre = 'TRABAJADOR POR CUENTA PROPIA'

Where fami_ocupa_padre = '9';

UPDATE saber11_2012_2016_transformado
SET fami_ocup_padre = 'HOGAR'

Where fami_ocupa_padre = '10';

UPDATE saber11_2012_2016_transformado
SET fami_ocup_padre = 'PENSIONADO'

Where fami_ocupa_padre = '11';

UPDATE saber11_2012_2016_transformado
SET fami_ocup_padre = 'OTRA ACTIVIDAD U OCUPACION'

Where fami_ocupa_padre = '12';

42. SCRIPTS:

```
UPDATE saber11_2012_2016_transformado
SET fami_nivel_sisben = 'NIVEL 1'
Where fami_nivel_sisben = '1';
```

```
UPDATE saber11_2012_2016_transformado
SET fami_nivel_sisben = 'NIVEL 2'
Where fami_nivel_sisben = '2';
```

```
UPDATE saber11_2012_2016_transformado
SET fami_nivel_sisben = 'NIVEL 3'
Where fami_nivel_sisben = '3';
```

```
UPDATE saber11_2012_2016_transformado
SET fami_nivel_sisben = 'ESTA CLASIFICADA EN OTRO NIVEL DE SISBEN'
Where fami_nivel_sisben = '4';
```

```
UPDATE saber11_2012_2016_transformado
SET fami_nivel_sisben = 'NO ESTA CLASIFICADA POR EL SISBEN'
Where fami_nivel_sisben = '5';
```

43. SCRIPTS:

```
ALTER TABLE saber11_2012_2016_transformado ADD eco_condicion_vive
text;
```

44. SCRIPTS:

$i = (\text{personas habitando una vivienda}) / (\text{numero de dormitorios en la vivienda})$
generalmente se aceptan los valores

- Hasta 2.4 -> sin hacinamiento
- De 2.5 a 4.9 -> hacinamiento medio
- Mas de 4.9 -> hacinamiento critico

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN
fami_cuartos_hogar TYPE double USING (trim(fami_cuartos_hogar):: double);
```

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN
fami_personas_hogar TYPE double USING (trim(fami_personas_hogar)::
double);
```

```
ALTER TABLE saber11_2012_2016_transformado ADD ind double;
```

```
UPDATE saber11_2012_2016_transformado
```

```
SET ind = (fami_personas_hogar)/( fami_cuartos_hogar);
```

```
UPDATE saber11_2012_2016_transformado  
SET eco_condicion_vive = 'SIN HACINAMIENTO'  
Where ind =< 2.4 ;
```

```
UPDATE saber11_2012_2016_transformado  
SET eco_condicion_vive = 'HACINAMIENTO MEDIO'  
Where ind => 2.5 AND ind<= 4.9 ;
```

```
UPDATE saber11_2012_2016_transformado  
SET eco_condicion_vive = 'HACINAMIENTO CRITICO'  
Where ind > 4.9 ;
```

45. SCRIPTS:

```
ALTER TABLE saber11_2012_2016_transformado ADD  
eco_condicion_vivienda text;
```

46. SCRIPTS:

```
UPDATE saber11_2012_2016_transformado  
SET eco_condicion_vivienda = 'MALA'  
WHERE fami_pisoshogar = '1';
```

```
UPDATE saber11_2012_2016_transformado  
SET eco_condicion_vivienda = 'REGULAR'  
WHERE fami_pisoshogar = '2';
```

```
UPDATE saber11_2012_2016_transformado  
SET eco_condicion_vivienda = 'REGULAR'  
WHERE fami_pisoshogar = '3';
```

```
UPDATE saber11_2012_2016_transformado  
SET eco_condicion_vivienda = 'BUENA'  
WHERE fami_pisoshogar = '4';
```

```
UPDATE saber11_2012_2016_transformado  
SET eco_condicion_vivienda = 'BUENA'  
WHERE fami_pisoshogar = '5';
```

47. SCRIPTS:

```
ALTER TABLE saber11_2012_2016_transformado  
ADD eco_condicion_tic text;
```

48. SCRIPTS:

```
ALTER TABLE saber11_2012_2016_transformado ADD ind_tic int;
```

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN  
fami_telefono_fijo TYPE integer USING (trim(fami_telefono_fijo)::integer);
```

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN  
fami_celular TYPE integer USING (trim(fami_celular)::integer);
```

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN  
fami_internet TYPE integer USING (trim(fami_internet)::integer);
```

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN  
fami_servicio_television TYPE integer USING  
(trim(fami_servicio_television)::integer);
```

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN  
fami_computador TYPE integer USING (trim(fami_computador)::integer);
```

```
UPDATE saber11_2012_2016_transformado  
SET ind_tic = (fami_telefono_fijo)+(fami_celular)+(fami_internet)+  
(fami_servicio_television)+(fami_computador);
```

```
UPDATE saber11_2012_2016_transformado  
SET eco_condicion_tic = 'BUENA'  
WHERE int_tic= '5' OR int_tic= '4';
```

```
UPDATE saber11_2012_2016_transformado  
SET eco_condicion_tic = 'REGULAR'  
WHERE int_tic= '2' OR int_tic= '3';
```

```
UPDATE saber11_2012_2016_transformado  
SET eco_condicion_tic = 'MALA'  
WHERE int_tic = '0' OR int_tic = '1';
```

49. SCRIPTS:

```
ALTER TABLE saber11_2012_2016_transformado  
ADD eco_condicion_electrodomesticos text;
```

50. SCRIPTS:

```
ALTER TABLE saber11_2012_2016_transformado ADD ind_elec int;
```

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN
fami_nevera TYPE integer USING (trim(fami_nevera)::integer);
```

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN
fami_horno TYPE integer USING (trim(fami_horno)::integer);
```

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN
fami_microondas TYPE integer USING (trim(fami_microondas)::integer);
```

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN
fami_lavadora TYPE integer USING (trim(fami_lavadora)::integer);
```

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN fami_dvd
TYPE integer USING (trim(fami_dvd)::integer);
```

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN
fami_televisor TYPE integer USING (trim(fami_televisor)::integer);
```

```
UPDATE saber11_2012_2016_transformado
SET ind_elec = (fami_nevera)+( fami_horno)+( fami_microondas)+
(fami_lavadora)+( fami_dvd);
```

```
UPDATE saber11_2012_2016_transformado
SET eco_condicion_electrodomesticos = 'BUENA'
WHERE int_elec= '5' OR int_elec= '4 OR int_elec= '6';
```

```
UPDATE saber11_2012_2016_transformado
SET eco_condicion_electrodomesticos = 'REGULAR'
WHERE int_elec = '2' OR int_elec= '3';
```

```
UPDATE saber11_2012_2016_transformado
SET eco_condicion_electrodomesticos = 'MALA'
WHERE int_elec = '0' OR int_elec = '1';
```

51. SCRIPTS:

```
UPDATE saber11_2012_2016_transformado
SET fami_ingreso_fmiliar_mensual= 'MENOS DE 1 SM'
WHERE fami_ingreso_fmiliar_mensual = '1' ;
```

```
UPDATE saber11_2012_2016_transformado
SET fami_ingreso_fmiliar_mensual= 'ENTRE 1 Y MENOS DE 2 SM'
```

```
WHERE fami_ingreso_fmiliar_mensual = '2' ;
```

```
UPDATE saber11_2012_2016_transformado  
SET fami_ingreso_fmiliar_mensual= 'ENTRE 2 Y MENOS DE 3 SM'  
WHERE fami_ingreso_fmiliar_mensual = '3' ;
```

```
UPDATE saber11_2012_2016_transformado  
SET fami_ingreso_fmiliar_mensual= 'ENTRE 3 Y MENOS DE 5 SM'  
WHERE fami_ingreso_fmiliar_mensual = '4' ;
```

```
UPDATE saber11_2012_2016_transformado  
SET fami_ingreso_fmiliar_mensual= 'ENTRE 5 Y MENOS DE 7 SM'  
WHERE fami_ingreso_fmiliar_mensual = '5' ;
```

```
UPDATE saber11_2012_2016_transformado  
SET fami_ingreso_fmiliar_mensual= 'ENTRE 7 Y MENOS DE 10 SM'  
WHERE fami_ingreso_fmiliar_mensual = '6' ;
```

```
UPDATE saber11_2012_2016_transformado  
SET fami_ingreso_fmiliar_mensual= '10 O MAS SM'  
WHERE fami_ingreso_fmiliar_mensual = '7' ;
```

52. SCRIPTS:

```
ALTER TABLE saber11_2012_2016_transformado ADD pun_global int;
```

53. SCRIPTS:

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN  
punt_biologia TYPE integer USING (trim(punt_biologia)::integer);
```

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN  
punt_c_naturales TYPE integer USING (trim(punt_c_naturales)::integer);
```

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN  
punt_c_sociales TYPE integer USING (trim(punt_c_sociales)::integer);
```

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN  
punt_comp_ciudadana TYPE integer USING  
(trim(punt_comp_ciudadana)::integer);
```

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN  
punt_comp_flexible TYPE integer USING (trim(punt_comp_flexible)::integer);
```

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN
punt_filosofia TYPE integer USING (trim(punt_filosofia)::integer);
```

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN
punt_fisica TYPE integer USING (trim(punt_fisica)::integer);
```

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN
punt_global TYPE integer USING (trim(punt_global)::integer);
```

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN
punt_ingles TYPE integer USING (trim(punt_ingles)::integer);
```

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN
punt_lectura_critica TYPE integer USING (trim(punt_lectura_critica)::integer);
```

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN
punt_lenguaje TYPE integer USING (trim(punt_lenguaje)::integer);
```

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN
punt_matematicas TYPE integer USING (trim(punt_matematicas)::integer);
```

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN
punt_quimica TYPE integer USING (trim(punt_quimica)::integer);
```

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN
punt_razona_cuant TYPE integer USING (trim(punt_razona_cuant)::integer);
```

```
ALTER TABLE saber11_2012_2016_transformado ALTER COLUMN
punt_sociales_ciudadanas TYPE integer USING
(trim(punt_sociales_ciudadanas)::integer);
```

```
UPDATE saber11_2012_2016_transformado
SET pun_global = (punt_biologia)+(punt_c_sociales)+(pun_comp_flexible)+
(punt_fisica)+(punt_filosofia)+(punt_ingles)+(punt_lenguaje)+(punt_matematica
s)+(punt_quimica)
```

```
Where periodo ='20121' OR periodo ='20122' OR periodo ='20131' OR periodo
='20132' OR periodo ='20141';
```

54. SCRIPTS:

```
ALTER TABLE saber11_2012_2016_transformado
ADD prom_rend int;
```

55. SCRIPTS:

```
UPDATE saber11_2012_2016_transformado
SET prom_rend = (pun_global)/9;
```

```
UPDATE saber11_2012_2016_transformado
SET prom_rend = (punt_global)/5
WHERE periodo=20161 OR periodo=20162 OR periodo=20151 OR
periodo=20152 OR periodo=20142;
```

56. SCRIPTS:

```
ALTER TABLE saber11_2012_2016_transformado
ADD rendimiento text;
```

57. SCRIPTS:

```
UPDATE saber11_2012_2016_transformado
SET rendimiento = 'EXCELENTE'
Where prom_rend > 80 ;
```

```
UPDATE saber11_2012_2016_transformado
SET rendimiento = 'BUENO'
Where prom_rend > 60 AND prom_rend <= 80 ;
```

```
UPDATE saber11_2012_2016_transformado
SET rendimiento = 'REGULAR'
Where prom_rend > 40 AND prom_rend <= 60;
```

```
UPDATE saber11_2012_2016_transformado
SET rendimiento = 'DEFICIENTE'
Where prom_rend <= 40 ;
```

58. SCRIPTS:

```
CREATE TABLE saber11_2012_2016_final
AS SELECT * FROM saber11_2012_2016_transformado;
```

59. SCRIPTS:

```
alter table saber11_2012_2016_final
drop column DECIL_C_NATURALES;
```

60. SCRIPTS:

```
alter table saber11_2012_2016_final
drop column DECIL_COMP_CIUADADANA;
```

61. SCRIPTS:
alter table saber11_2012_2016_final
drop column DECIL_INGLES;

62. SCRIPTS:
alter table saber11_2012_2016_final
drop column DECIL_LLECTURA_CRITICA;

63. SCRIPTS:
alter table saber11_2012_2016_final
drop column DECIL_MATEMATICAS;

64. SCRIPTS:
alter table saber11_2012_2016_final
drop column DECIL_RAZONA_CUANT;

65. SCRIPTS:
alter table saber11_2012_2016_final
drop column DECIL_SOCIALES_CIUADANAS;

66. SCRIPTS:
alter table saber11_2012_2016_final
drop column ESTU_CUANTO_FALTO_COLEGIO;

67. SCRIPTS:
alter table saber11_2012_2016_final
drop column DECIL_C_NATURALES;

68. SCRIPTS:
alter table saber11_2012_2016_final
drop column DESEMP_LLECTURA_CRITICA;

69. SCRIPTS:
alter table saber11_2012_2016_final
drop column DESEMP_MATEMATICAS;

70. SCRIPTS:
alter table saber11_2012_2016_final
drop column DESEMP_SOCIALES_CIUADANAS;

71. SCRIPTS:
alter table saber11_2012_2016_final

drop column PERCENTIL_C_NATURALES;

72. SCRIPTS:

alter table saber11_2012_2016_final
drop column PERCENTIL_GLOBAL;

73. SCRIPTS:

alter table saber11_2012_2016_final
drop column PERCENTIL_INGLES;

74. SCRIPTS:

alter table saber11_2012_2016_final
drop column PERCENTIL_LLECTURA_CRITICA;

75. SCRIPTS:

alter table saber11_2012_2016_final
drop column PERCENTIL_MATEMATICAS;

76. SCRIPTS:

alter table saber11_2012_2016_final
drop column PERCENTIL_SICIALES_CIUADANAS;

77. SCRIPTS:

alter table saber11_2012_2016_final
drop column FAMI_TELEVISOR;