

**UN MERCADO DE DATOS PARA EL ANÁLISIS MULTIDIMENSIONAL DE LAS
PRUEBAS SABER 9 DE LAS INSTITUCIONES EDUCATIVAS DE LOS
MUNICIPIOS DE LA SUBREGIÓN DE OBANDO DEL DEPARTAMENTO DE
NARIÑO**



Universidad de **Nariño**

**ANDRÉS FELIPE ACOSTA YAMPUEZÁN
JORGE ALBERTO BARAHONA CADENA**

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

**UN MERCADO DE DATOS PARA EL ANÁLISIS MULTIDIMENSIONAL DE LAS
PRUEBAS SABER 9 DE LAS INSTITUCIONES EDUCATIVAS DE LOS
MUNICIPIOS DE LA SUBREGIÓN DE OBANDO DEL DEPARTAMENTO DE
NARIÑO**

**ANDRÉS FELIPE ACOSTA YAMPUEZÁN
JORGE ALBERTO BARAHONA CADENA**

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

**DIRECTOR
PhD. RICARDO TIMARÁN PEREIRA**

**UNIVERSIDAD DE NARIÑO
FACULTAD DE INGENIERÍA
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 de los autores.”

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 del presidente del jurado

Firma del jurado

Firma del jurado

Firma del asesor

Firma del co-asesor

San Juan de Pasto, agosto del 2018

AGRADECIMIENTOS

A Dios, por colmarnos de bendiciones, paciencia y sabiduría a lo largo del desarrollo de este proyecto y de nuestra carrera universitaria.

A la Universidad de Nariño, a los profesores y de antemano un agradecimiento especial al Ingeniero Ricardo Timaran Pereira quien, con su conocimiento y su experiencia en el campo investigativo, apporto al desarrollo y conclusión de este proyecto.

Finalmente, a nuestros familiares y amigos por su apoyo incondicional que nos brindaron.

RESUMEN

En este documento se presentan los resultados del proyecto de investigación que tuvo como objetivo construir un mercado de datos para almacenar la información histórica de los estudiantes del grado 9º de las Instituciones Educativas de los municipios de la subregión de Obando del departamento de Nariño (Colombia) que presentaron las pruebas saber 9º en los años 2014, 2015 y 2016. Se utilizó como metodología para diseñar el modelo multidimensional del mercado de datos, 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 resultados. Este mercado de datos brinda información limpia, transformada y de calidad que ayudara a los directivos de las Instituciones Educativas de los municipios de la subregión de Obando y a las Secretarías de Educación municipales, a la toma de decisiones en lo relacionado con la formulación de estrategias para el mejoramiento de la calidad educativa de las instituciones de esta subregión.

Palabras clave: Mercado de Datos, Análisis Multidimensional, Pruebas Saber 9º, Instituciones Educativas, Subregión de Obando, Departamento de Nariño

ABSTRACT

In this document the results of research project aimed to build a data mart to store the historical information of the students of the Educational Institutions of the municipalities of the Obando subregion of the department of Nariño Colombia who presented the tests Saber 9th between the years 2014 to 2016, are presented. As methodology was used the proposed by Kimball to design the multidimensional model of the data mart.

The data mart was integrated into the Pentaho free business intelligence tool to perform multidimensional analysis and visualization of results. This data mart provides clean, transformed and quality information that will help the managers of the Educational Institutions of the municipalities of the Obando subregion and the Municipal Secretaries of Education, to make decisions regarding the formulation of strategies for the improvement of the educational quality of the institutions this subregion.

Keywords: Data mart, Multidimensional Analysis, Tests Saber 9th, Educational Institutions, Obando Subregion, Department of Nariño.

CONTENIDO

	Pág.
1. INTRODUCCIÓN	14
Fundamentos teóricos	20
Business Intelligence, BI	20
Beneficios de Business Intelligence	20
Data Warehouse, DW	21
Arquitectura data warehouse.....	21
Componentes básicos data warehouse	23
Representación data warehouse.....	25
Procesos ETL: Extracción, Transformación y Carga de datos	25
Proceso de Extracción.	26
Proceso de Transformación	27
Proceso de Carga	27
Técnicas de explotación data warehouse.....	27
Sistema OLAP:.....	28
Operaciones OLAP	29
Sistema MOLAP	30
Sistema ROLAP	31
Mercado de datos (Datamart).....	33
Data warehouse y mercado de datos.....	34
Diseño de un mercado de datos.....	35
¿Qué metodología adoptar?.....	35
Metodología Kimball.....	36
Pruebas Saber 9°	38
¿Cuál es el objetivo de Saber 9°?.....	39
¿A quiénes se evalúa?	39
¿Qué se evalúa?.....	39
Estructura y alineación de las pruebas saber 9°	40
La Ex Provincia de Obando.....	41
Aspectos demográficos.	42

Metodología.....	43
Planificación del proyecto.....	43
Análisis de requerimientos	43
Modelado dimensional	44
Diseño Físico.....	45
Diseño del subsistema de Extracción, Transformación y Carga (ETL)	45
Especificación y desarrollo de aplicaciones de BI	46
Reportes Estándar	46
Aplicaciones analíticas.	46
2. IMPLEMENTACIÓN	47
2.1. Análisis de Requerimientos	47
2.2. Modelado dimensional	49
2.3. Diseño del sistema de Extracción, Transformación y Carga (ETL)	51
2.3.1. Fase de Extracción.....	51
2.3.2. Fase de Transformación.....	65
2.3.3. Fase de Carga.....	65
2.4. Implementación del mercado de datos en PostgreSQL	74
2.5. Diccionario del mercado de datos	76
2.6. Integración del mercado de datos con Pentaho	79
2.6.1. Cubo de datos.	79
2.6.2. Operaciones OLAP	81
2.7. Especificación y desarrollo de aplicaciones de BI	85
3. RESULTADOS Y DISCUSIÓN.....	88
4. CONCLUSIONES Y RECOMENDACIONES	123
BIBLIOGRAFÍA	125
ANEXOS	128

LISTA DE TABLAS

	Pág.
Tabla 1. Ventajas y desventajas sistemas MOLAP.....	31
Tabla 2. Ventajas y desventajas sistemas ROLAP.....	32
Tabla 3. Diferencias de un datawarehouse con un sistema tradicional.	32
Tabla 4. Data warehouse vs Mercado de datos.....	34
Tabla 5. Procesos de las competencias genéricas.....	40
Tabla 6. Componentes de las competencias genéricas.....	40
Tabla 7. Población municipios de la Ex Provincia de Obando.....	42
Tabla 8. Consolidado de las posibles combinaciones entre competencias genéricas	47
Tabla 9. Dimensiones y granularidad del mercado de datos.....	49
Tabla 10. Valores Plausibles.....	52
Tabla 11. Establecimientos.....	54
Tabla 12. Entidades.....	54
Tabla 13. Municipio.....	55
Tabla 14. Sedes.....	55
Tabla 15. Análisis de la calidad de valores plausibles.....	56
Tabla 16. Análisis de la calidad de establecimientos.....	60
Tabla 17. Análisis de la calidad de Ente Territorial.....	61
Tabla 18. Análisis de la calidad de Municipio.....	63
Tabla 19. Análisis de la calidad de Sedes.....	63
Tabla 20. Diccionario de datos de la dimensión Estudiante.....	76
Tabla 21. Diccionario de datos de la dimensión Sedes.....	76
Tabla 22. Diccionario de datos de la dimensión Establecimientos.....	77
Tabla 23. Diccionario de datos de la dimensión Tipo Pruebas.....	77
Tabla 24. Diccionario de datos de la dimensión Ente territorial.....	78
Tabla 25. Diccionario de datos de la dimensión Municipio.....	78
Tabla 26. Diccionario de datos de la dimensión Tiempo.....	78
Tabla 27. Diccionario de datos de la tabla de hechos.....	79

LISTA DE FIGURAS

	Pág.
Figura 1. Estructura básica de un data warehouse.....	22
Figura 2. Procesos ETL: Extracción, transformación y carga de datos.....	26
Figura 3. Representación sistema MOLAP (Multidimensional Online Analytical Processing).....	30
Figura 4. Representación sistema ROLAP (Relational Online Analytical Processing).	31
Figura 5. Esquema de un mercado de datos.	33
Figura 6. Fases de la metodología Kimball.	37
Figura 7. Diagrama de flujo del proceso dimensional de Kimball.....	44
Figura 8. Esquema en estrella.	50
Figura 9. Proceso ETL automatizado en la herramienta Pentaho Data Integration (PDI) el cual fue utilizado en la carga de datos de la dimensión ente territorial.	66
Figura 10. Proceso ETL automatizado en la herramienta Pentaho Data Integration (PDI) el cual fue utilizado en la carga de datos de la dimensión establecimientos.	67
Figura 11. Proceso ETL automatizado en la herramienta Pentaho Data Integration (PDI) el cual fue utilizado en la carga de datos de la dimensión municipios.	68
Figura 12. Proceso ETL automatizado en la herramienta Pentaho Data Integration (PDI) el cual fue utilizado en la carga de datos de la dimensión sedes.	69
Figura 13. Proceso ETL automatizado en la herramienta Pentaho Data Integration (PDI) el cual fue utilizado en la carga de datos de la dimensión tiempo.	70
Figura 14. Proceso ETL automatizado en la herramienta Pentaho Data Integration (PDI) el cual fue utilizado en la carga de datos de la dimensión tipo prueba.	70
Figura 15. Proceso ETL automatizado en la herramienta Pentaho Data Integration (PDI) el cual fue utilizado en la carga de datos de la tabla de hechos.	71
Figura 16. Proceso ETL automatizado en la herramienta Pentaho Data Integration (PDI) el cual fue utilizado en la transformación de plausibles.	73
Figura 17. Implementación del Modelo Multidimensional tipo estrella en PostgreSQL.	75
Figura 18. Configuración del cubo para las consultas OLAP en Schema Workbench.	80
Figura 19. Promedio por municipio y por competencias genéricas.	81

Figura 20. Promedio por establecimiento educativo, ente territorial y año realización de la prueba.	82
Figura 21. Promedio por municipio en lenguaje y ciencias.	83
Figura 22. Promedio por municipio en lenguaje y competencias.	83
Figura 23. Promedio por municipio en lenguaje y matemáticas.	84
Figura 24. Promedio por municipio en matemáticas y ciencias.	84
Figura 25. Promedio por municipio en matemáticas y competencias.	85
Figura 26. Pantalla de inicio Pentaho.	86
Figura 27. Creación de reporte en Pentaho.	86
Figura 28. Sentencia SQL para diseño de cubos.	87
Figura 29. Número de estudiantes que presentaron las Pruebas Saber 9° entre 2014 y 2016 clasificados por municipio.	88
Figura 30. Puntaje promedio de los resultados de lenguaje y ciencias por municipio en la subregión de Obando.	89
Figura 31. Puntaje promedio de los resultados de lenguaje y competencias por municipio en la subregión de Obando.	90
Figura 32. Puntaje promedio de los resultados de lenguaje y matemáticas por municipio en la subregión de Obando.	91
Figura 33. Puntaje promedio de los resultados de matemáticas y ciencias por municipio en la subregión de Obando.	92
Figura 34. Puntaje promedio de los resultados de matemáticas y competencias por municipio en la subregión de Obando.	93
Figura 35. Número de estudiantes por municipio y género que presentaron las pruebas saber 9° comprendido entre los años 2014 a 2016.	94
Figura 36. Puntaje promedio de los resultados por municipio, clasificado por género en las áreas de lenguaje y ciencias.	95
Figura 37. Puntaje promedio de los resultados por municipio, clasificado por género en las áreas de lenguaje y ciencias.	96
Figura 38. Puntaje promedio de los resultados por municipio, clasificado por género en las áreas de lenguaje y matemáticas.	97
Figura 39. Puntaje promedio de los resultados por municipio, clasificado por género en las áreas de matemáticas y ciencias.	98
Figura 40. Puntaje promedio de los resultados por municipio, clasificado por género en las áreas de matemáticas y competencias.	99
Figura 41. Puntaje promedio de la prueba de lenguaje y ciencias clasificada por género.	100
Figura 42. Porcentaje de estudiantes que presentaron la prueba de lenguaje y ciencias, clasificada por género.	101
Figura 43. Puntaje promedio de la prueba de lenguaje y competencias clasificada por género.	102
Figura 44. Porcentaje de estudiantes que presentaron la prueba de lenguaje y competencias, clasificada por género.	102
Figura 45. Puntaje promedio de la prueba de lenguaje y matemáticas clasificada por género.	103

Figura 46. Porcentaje de estudiantes que presentaron la prueba de lenguaje y matemáticas, clasificada por género.	104
Figura 47. Porcentaje de estudiantes que presentaron la prueba de lenguaje y matemáticas, clasificada por género.	104
Figura 48. Porcentaje de estudiantes que presentaron la prueba de matemáticas y ciencias, clasificada por género.	105
Figura 49. Puntaje promedio de la prueba de matemáticas y competencias clasificada por género.	106
Figura 50. Porcentaje de estudiantes que presentaron la prueba de matemáticas y competencias, clasificada por género.	106
Figura 51. Puntaje promedio de los resultados obtenidos por año en el área de lenguaje y ciencias.	107
Figura 52. Puntaje promedio de los resultados obtenidos por año en el área de lenguaje y competencias.	108
Figura 53. Puntaje promedio de los resultados obtenidos por año en el área de lenguaje y matemáticas.	109
Figura 54. Puntaje promedio de los resultados obtenidos por año en el área de matemáticas y ciencias.	110
Figura 55. Puntaje promedio de los resultados obtenidos por año en el área de matemáticas y competencias.	111
Figura 56. Puntaje promedio de los resultados en lenguaje y ciencias, clasificados por tipo de establecimiento.	112
Figura 57. Puntaje promedio de los resultados en lenguaje y competencias, clasificados por tipo de establecimiento.	113
Figura 58. Puntaje promedio de los resultados en lenguaje y matemáticas, clasificados por tipo de establecimiento.	114
Figura 59. Puntaje promedio de los resultados en matemáticas y ciencias, clasificados por tipo de establecimiento.	115
Figura 60. Puntaje promedio de los resultados en matemáticas y competencias, clasificados por tipo de establecimiento.	116
Figura 61. Puntaje promedio de los resultados en lenguaje y ciencias, clasificados por zona.	117
Figura 62. Puntaje promedio de los resultados en lenguaje y competencias, clasificados por zona.	118
Figura 63. Puntaje promedio de los resultados en lenguaje y matemáticas, clasificados por zona.	119
Figura 64. Puntaje promedio de los resultados en matemáticas y ciencias, clasificados por zona.	120
Figura 65. Puntaje promedio de los resultados en matemáticas y competencias, clasificados por zona.	121

INTRODUCCIÓN

El propósito de la política educativa es garantizar que todos los estudiantes, independientemente de su procedencia y contexto socioeconómico y cultural en el que viven, reciban en la escuela una educación de alta calidad, que contribuya al desarrollo de las competencias necesarias para vivir, convivir, ser productivos en todos los ámbitos y seguir aprendiendo a lo largo de la vida.¹

Por tal razón, el Ministerio de Educación Nacional (MEN) ha venido buscando estrategias desde comienzos de los años 90, para cumplir dicho propósito, y una de estas estrategias es la aplicación de evaluaciones externas, fundamentadas en la medición de logros mediante las pruebas estructuradas por competencias, las cuales son esenciales para que todos los colegios conozcan sus avances y puedan generar los cambios para mejorar el sistema educativo.²

La ley 1324 le confiere al Instituto Colombiano para la Evaluación de la Educación (ICFES), la misión de evaluar, y hacer seguimiento a la educación, mediante pruebas estandarizadas, saber 3°, 5°, y 9°.

Con estas pruebas es posible evaluar lo que los estudiantes saben y saben hacer, con lo que aprenden durante su trayectoria escolar y a partir de sus resultados, identificar los avances, así como las deficiencias o debilidades sobre las cuales es preciso tomar medidas para mejorar. Estas evaluaciones son las que se realizan internamente mediante la valoración permanente de los aprendizajes de los estudiantes, o externamente a través de pruebas estandarizadas.³

Específicamente en las instituciones educativas de la subregión de Obando, el análisis de la Pruebas Saber 9°, se realiza de manera aislada, en cada establecimiento sin tener en cuenta el resultado a nivel regional.

De acuerdo a lo mencionado anteriormente, se construyó un mercado de datos para las Instituciones Educativas de la subregión de Obando, que almacena datos

¹ Ver Lineamientos generales Saber 2009 5° y 9°

² Ver Marco de factores asociados Saber 3°, 5° y 9° 2016

³ Ver Informe técnico Saber 5° y 9° 2009

históricos, limpios y transformados de los estudiantes que presentaron las pruebas saber 9° entre 2014 y 2016.

Esta información permite a las instituciones educativas y secretarías de educación, la generación de indicadores que soporte la toma de decisiones y oriente la generación de políticas y estrategias para el mejoramiento de la calidad educativa.

La metodología que se utilizó para la construcción del mercado de datos, es la que propone Ralph Kimball, la cual se basa en lo que se denomina Ciclo de vida dimensional del negocio (Business Dimensional Lifecycle), considerada una de las técnicas favoritas a la hora de construir sistemas data warehouse y Business Intelligence, cuyos resultados sirven para el apoyo a la toma de decisiones.

Planteamiento del Problema

Una de las grandes problemáticas en Colombia desde tiempos atrás hasta la actualidad es la calidad de la educación, reflejada específicamente en los resultados de las Pruebas Saber, con desempeños bajos e insuficientes.

El Sistema Nacional de Evaluación de la Calidad de la educación fue creado por la Ley General de educación para evaluar periódicamente a todos los actores y componentes del servicio educativo, entre los cuales se incluyen los resultados de aprendizaje.

En cumplimiento de este mandato, desde los primeros años de la década de los 90 se inició el desarrollo de SABER, concebido como una evaluación nacional de calidad educativa y se aplicaron sucesivamente pruebas a muestras de estudiantes de determinados grados y en algunas áreas del conocimiento, con el propósito de medir sus logros. Los resultados de estas primeras evaluaciones proporcionan información valiosa para orientar la formulación de políticas y programas para apoyar el mejoramiento de la calidad educativa del país.⁴

El carácter muestral de estas primeras mediciones no permitía contar con información a un nivel crucial para la toma de decisiones dentro del sistema: el de los establecimientos educativos. Las evaluaciones censales trienales, obligatorias desde la expedición de la ley 715 de 2001, hacen posible que todas las instituciones educativas dispongan de información confiable acerca de, lo que sus estudiantes saben y saben hacer y, a partir de la misma, adelantar acciones específicas para superar las dificultades encontradas.⁵

⁴ Ver Alineación del examen saber 11

⁵ Ver Saber 5° y 9° Síntesis de resultados de factores asociados

Particularmente, en la subregión de Obando, la cual abarca los municipios de Aldana, Contadero, Córdoba, Cuaspud, Cumbal, Funes, Guachucal, Gualmatán, Iles, Ipiales, Potosí, Puerres y Pupiales, las Instituciones Educativas de dichos municipios presentan un bajo rendimiento en el desarrollo de las pruebas Saber 9°. Razón por la cual se encontró la necesidad de construir un mercado de datos que almacene los datos históricos, limpios y transformados de los estudiantes que presentaron las pruebas Saber 9° en el periodo comprendido durante los años 2014, 2015 y 2016, siendo este, un sistema orientado a la consulta, permitiendo así, arrojar información relevante como estrategia para la toma de decisiones y mejoramiento de calidad de educación en las Instituciones Educativas de la subregión de Obando.

Justificación

Este proyecto de investigación, tuvo como objetivo construir un mercado de datos que almacene la información histórica de los estudiantes de las Instituciones Educativas de la subregión de Obando en el departamento de Nariño, que presentaron las pruebas Saber 9° entre los años 2014 y 2016. Este mercado, gestionado por una herramienta de Inteligencia de Negocios de software libre como lo es Pentaho, permitió 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 Educativas de la subregión de Obando.

Este mercado de datos de las pruebas Saber 9°, diseñado bajo un modelo multidimensional, que almacena información limpia y transformada, permitió la distribución de los resultados del análisis en múltiples formatos y de forma inmediata.

Este proyecto buscó generar información confiable y de calidad acerca de los resultados de las pruebas Saber 9° para soportar la toma de decisiones de los directivos de las Instituciones Educativas, de las secretarías de educación en la subregión de Obando e incluso del 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.

Objetivos

Objetivo General. El objetivo general de este proyecto de investigación fue, diseñar y construir un mercado de datos, que almacene datos históricos, limpios y transformados de los resultados obtenidos de los estudiantes de las Instituciones Educativas de la subregión de Obando en el departamento de Nariño, que presentaron las pruebas Saber 9° entre los años 2014, 2015 y 2016, permitiendo así, generar 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

Los objetivos específicos planteados en esta investigación para lograr el objetivo general fueron:

- Apropiar el conocimiento sobre las competencias evaluadas en las pruebas Saber 9 y la herramienta (software) de inteligencia de negocios, Pentaho.
- Seleccionar e integrar los datos de las bases de datos del ICFES de las pruebas Saber 9° del proyecto.
- Aplicar técnicas de limpieza y transformación de datos al repositorio temporal de datos de las pruebas Saber 9°.
- Diseñar e implementar el modelo multidimensional en estrella de los datos de las pruebas Saber 9° en un Sistema Gestor de Base de Datos Relacional (SGBDR).
- Analizar y evaluar los informes y reportes generados, por diferentes dimensiones, con la herramienta Pentaho acerca de las pruebas Saber 9°.

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 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. Reconociendo la importancia de esta información, se pudo obtener resultados estadísticos 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.

Organización de este documento

Este documento se organiza en 3 capítulos, que se describen a continuación.

En el capítulo 1, se describe los fundamentos teóricos en el cual se desarrolla esta investigación.

En el capítulo 2, se da a conocer la implementación de la metodología utilizada en la construcción del mercado de datos.

En el capítulo 3, se presenta los resultados, que se obtuvo de la construcción final del mercado de datos.

Fundamentos teóricos

Business Intelligence, BI

Business Intelligence, (Inteligencia de negocios), es el proceso de obtener información sobre el negocio a partir de los datos existentes. En la era de la información, las organizaciones tienen a su disposición cantidades de datos recolectados en sistemas transaccionales los cuales son esenciales para la operación del negocio⁶

Disponer de datos no es lo mismo que disponer de información. Los datos se convierten en información cuando se pueden utilizar para responder a cuestiones del negocio, de tal manera que se pueda comprender mejor el funcionamiento del mismo.

Para Espiñera, Sheldon y asociados, la inteligencia de negocios incluye una amplia categoría de metodologías, aplicaciones y tecnologías que permiten reunir, acceder, transformar y analizar los datos, transacciones e información no estructurada (interna y externa), con el propósito de ayudar a los usuarios de una compañía a tomar mejores decisiones de negocio.

Por otra parte, en el libro Business Intelligence⁷, menciona que el objetivo básico de inteligencia de negocios es apoyar de forma sostenible y continuada a las organizaciones para mejorar su competitividad, facilitando la información necesaria para la toma de decisiones.

Mientras que Dresner⁸, popularizo Business Intelligence o BI como un término para describir un conjunto de conceptos y métodos que mejoraran la toma de decisiones, utilizando información sobre qué había sucedido (hechos).

Beneficios de Business Intelligence. Los principales beneficios que se puede obtener al implementar una solución de inteligencia de negocios en una organización son:

⁶ Ver Espiñera, Sheldon y Asociados. La inteligencia de negocios (Business Intelligence)

⁷ Ver Lluís, Josep. Business Intelligence: Competir con información

⁸ Ver Dresner, H. Glosario términos

- Respuestas inmediatas a preguntas del negocio, que son básicas para la toma de decisiones.
- Integración de datos entre los diferentes sistemas de información existentes en la organización.
- Permite lograr una visión del futuro a través del análisis de datos históricos.
- Dar la libertad al usuario de crear diferentes escenarios de análisis, sin la dependencia del área de tecnología.

Data Warehouse, DW

Hoy en día las empresas cuentan en su mayoría con sus procesos automatizados, manejando gran cantidad de datos en forma centralizada y manteniendo sus sistemas en línea.

Sin embargo, si la información no es sólida para ayudar y apoyar las decisiones, la automatización no tiene sentido. Esta necesidad de obtener información para una amplia variedad de individuos es la principal razón de negocios que conduce al concepto de data warehouse⁹.

El DW puede verse como una bodega donde están almacenados todos los datos los cuales están dirigidos a satisfacer las necesidades de información de los analistas y usuarios encargados de tomar decisiones.

Los almacenes de datos o data warehouse, generan base de datos tangibles con una perspectiva histórica, utilizando datos de múltiples fuentes que se fusionan en forma congruente.

Arquitectura data warehouse. La estructura básica de la arquitectura de un data warehouse incluye:

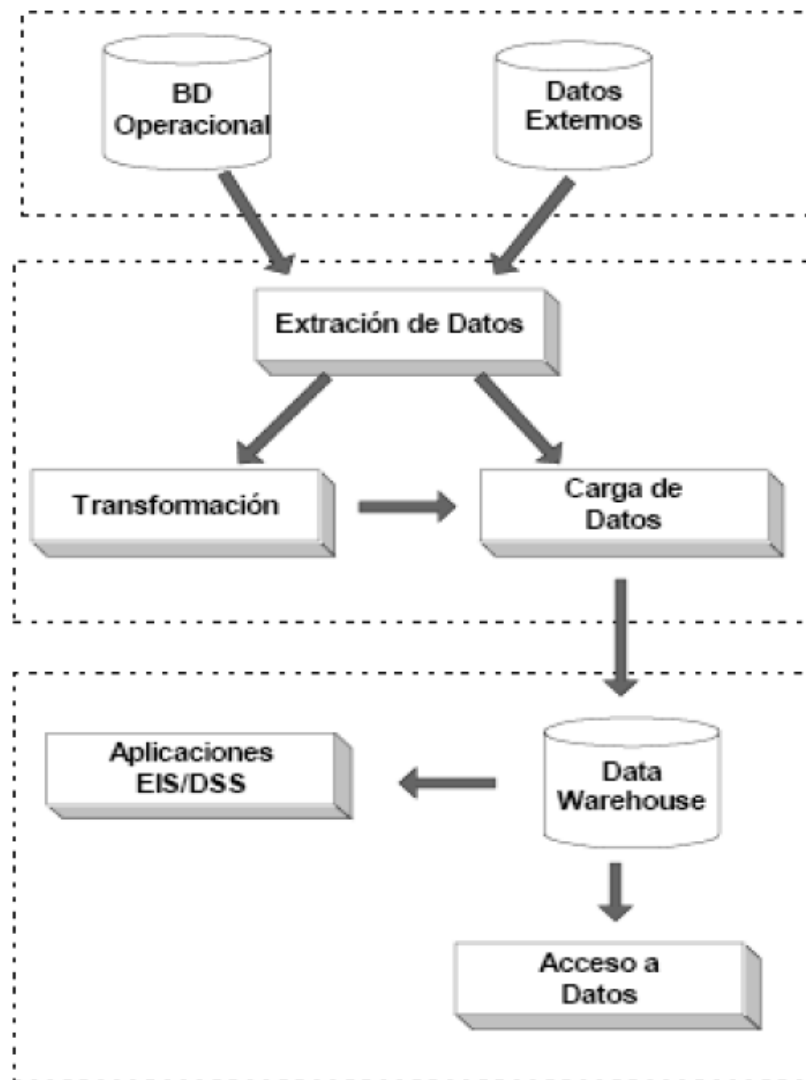
1. Datos Operacionales: el origen de los datos son los sistemas transaccionales internos de la organización como también datos externos a esta para poblar el componente de almacenamiento físico DW.
2. Extracción de datos: selección sistemática de datos operacionales utilizados para poblar el componente de almacenamiento físico DW.

⁹ Sperley, Eric. The Enterprise Data Warehouse: Planning, Building, and Implementation”

3. Transformación de datos: procesos para realizar cambios en los datos operacionales para reunir los objetivos de orientación a temas de integración.
4. Carga de datos: inserción sistemática de datos en el componente de almacenamiento físico DW.
5. Data warehouse: almacenamiento físico de datos de la arquitectura DW.
6. Herramientas de acceso al componente de almacenamiento físico DW: herramientas que proveen acceso a los datos, las cuales pueden ser específicas para visualización de bases multidimensionales almacenadas en data warehouse.

La figura 1 muestra la estructura básica de un data warehouse.

Figura 1. Estructura básica de un data warehouse.



Fuente:¹⁰

Componentes básicos data warehouse. A la hora de realizar un modelo dimensional del negocio se hace necesaria la presentación de los siguientes componentes:

- Dimensiones

Son las áreas temáticas, líneas o sujetos del negocio. Las dimensiones definidas como un grupo de uno o más atributos, separados y distintos unos de otros según el esquema de implementación elegido.

¹⁰ M. Reyes y P. Rosales, “Desarrollo de un datamart de información académica de estudiantes”

- Atributos

Es la agrupación de elementos o ítems dentro de una dimensión, los cuales representan categorías o clases de elementos que tienen el mismo nivel lógico dentro de una dimensión donde todos los elementos de un atributo se relacionan con otros atributos de la dimensión de la misma forma. La finalidad de los atributos es ver la información de cada dimensión a diferentes niveles de detalle y agrupar los datos para ser analizados.

- Relaciones

Son asociaciones lógicas de atributos dentro de una jerarquía definida por las instancias de los atributos y transitivas dentro de la jerarquía. Los atributos dentro de una dimensión están directamente relacionados unos con otros a través de los diferentes tipos de relaciones.

- Jerarquías

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.

- Indicadores

Es una variable dimensional unitaria, que correlaciona dos variables cualesquiera. A través de este concepto es posible relacionar diversas variables presentes en los procesos de la empresa.

La importancia de un indicador radica en la particularidad de informar al usuario el estado actual del sistema el cual permitirá tomar decisiones preventivas o correctivas de acuerdo a los resultados de las comparaciones entre el valor esperado y el valor obtenido.

- Tabla de hechos

Los hechos son los indicadores de negocio que dan sentido al análisis de las dimensiones. Las tablas de hechos incluyen los indicadores asociados a un proceso de negocio en concreto y las claves de las dimensiones que intervienen en dicho proceso, en el mínimo nivel de granularidad o detalle.

Representación data warehouse. Existen varias formas de representación o implementación del modelo dimensional dentro de las diferentes bases de datos que existen en el mercado. En el caso de contarse con un ambiente netamente relacional, es decir, una base de datos relacional sobre la cual se hará la implementación del data warehouse, se suele utilizar una representación denominada, esquema en estrella, la cual proporciona simpleza y flexibilidad para el uso de cualquier herramienta de usuario final.

Esquema en estrella

El modelo de esquema en estrella puede verse como una simple estrella en la cual existe una tabla central que contiene los hechos del negocio que se desean modelar, y múltiples tablas radiantes, llamadas dimensiones, conectadas a la central a través de las respectivas llaves primarias y foráneas. A diferencia de la estructura de otros esquemas de base de datos, un esquema en estrella contiene dimensiones desnormalizadas.

Los esquemas en estrella presentan las siguientes características:

- El modelo es fácil de entender para los usuarios.
- La llave primaria representa a cada una de las dimensiones.
- Las columnas que no son llaves foráneas son valores.
- Las tablas de hechos están usualmente normalizadas.
- Las dimensiones están completamente desnormalizadas.
- Proveen una respuesta rápida a las consultas.
- El rendimiento de las consultas es mejorado reduciendo las uniones entre tablas.
- Los usuarios pueden expresar consultas complejas.
- Son ambientes ideales para proveer información a través de múltiples herramientas de presentación.

Procesos ETL: Extracción, Transformación y Carga de datos

Los procesos extracción, transformación y carga de datos (ETL) que se muestran en la figura 2, son métodos para referirse al movimiento y transformación de datos que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y cargarlos en otra base de datos denominada mercado de datos o

data warehouse con el objeto de analizarlos y posteriormente influir en la toma de decisiones.

Figura 2. Procesos ETL: Extracción, transformación y carga de datos.



Fuente: ¿Qué es un data warehouse? (<http://www.dataprix.com>).

Proceso de Extracción. La primera etapa del proceso ETL consiste en identificar donde se almacenan los datos relevantes, es decir los sistemas o fuentes de origen, lo cual puede llegar a ser un desafío ya que el almacenamiento de los datos puede tener diversos orígenes, donde estos a la vez generan información en soporte magnético, pero también hay casos que se genera la información en formato físico (cupones, cuestionarios o formularios) que también deben ser extraídos.

Cada sistema separado puede usar una organización diferente de los datos o formatos distintos, donde los formatos de las fuentes pueden estar en base de datos relacionales o ficheros planos como también pueden estar en base de datos no relacionales u estructuras diferentes.

Una característica propia del proceso de extracción es la de analizar los datos extraídos, para verificar si los datos cumplen con los requisitos o la estructura que se esperaba, de no ser así los datos son rechazados. La extracción convierte los datos a un formato preparado para iniciar el proceso de transformación.

Para realizar de manera correcta el proceso de extracción se siguió los siguientes pasos:

- Extraer los datos desde los sistemas de origen.
- Analizar los datos extraídos obteniendo un control de los mismos
- Interpretar el resultado para verificar que los datos extraídos cumplen con los requisitos o estructura que se esperaba. Si no resulta así, los datos deberían ser desechados.
- Convertir los datos a un formato preparado para iniciar el proceso de transformación.

Proceso de Transformación. El proceso de transformación, requiere una serie de reglas de negocio o funciones sobre los datos extraídos para posteriormente comprobar su calidad, eliminar datos duplicados y cuando es posible corregir valores erróneos y completar valores vacíos, es decir, se transforman los datos siempre que sea posible para reducir errores de carga.

El objetivo de este proceso es recuperar los datos limpios y de alta calidad con el propósito de estructurar y resumir en los distintos modelos de análisis donde el resultado de este proceso es la obtención de datos limpios, consistentes y útiles. La transformación incluye: cambios de formato, sustitución de códigos, valores derivados y agregados, etc.

Proceso de Carga. En esta etapa, los datos procedentes del anterior proceso, transformación, son cargados en el sistema de destino ya sea en un data warehouse o un mercado de datos. Dependiendo de los requerimientos establecidos, este proceso puede abarcar una amplia variedad de acciones diferentes, como en el caso de algunas bases de datos donde se sobrescribe la información antigua con nuevos datos.

Los data warehouse mantienen un historial de los registros de manera que se pueda hacer un análisis de los mismos y disponer de un rastro de toda la historia de un valor a lo largo del tiempo.

Técnicas de explotación data warehouse

La explotación del data warehouse requiere de una herramienta óptima que permita la búsqueda y análisis de la información que se encuentra almacenada en el data warehouse y posteriormente contribuir a la toma de decisiones en la organización.

Los sistemas de soporte a la decisión usando tecnologías de data warehouse se conocen como sistemas OLAP (Online Analytical Processing) la cual permite al usuario organizar, visualizar, resumir y analizar la información de forma rápida, amigable e interactiva desde distintas perspectivas y facilitar la construcción de modelos cuantitativos complejos.

Sistema OLAP. Las herramientas OLAP tienen como objetivo fundamental soportar las consultas de los usuarios que analizan el negocio. Por lo general estos usuarios están familiarizados con el uso de las hojas de cálculo, las herramientas OLAP, básicamente realizan las mismas operaciones que podríamos realizar con una hoja de cálculo, pero además ofrecen:

- Trabajar con grandes volúmenes de datos.
- Están optimizadas semánticamente para manejar términos de negocio (dimensiones geográficas, dimensioe de tiempo, etc.)
- Están combinados con capacidades de generación de informes.
- Trabajan con bases de datos dimensionales.

Los sistemas OLAP son frecuentemente comparados con lo que se denominan base de datos estadísticas, una clase de bases de datos que permiten la definición, manipulación, elaboración y almacenamiento de datos multidimensionales pre calculados, la principal diferencia entre ambas radica en que las bases de datos estadísticas calculan los datos sobre otras bases de datos, mientras que las bases de datos OLAP representan los datos directamente.

Los modelos de datos OLAP, basados en datos organizados de manera multidimensional, típicamente categorizan los datos como dimensiones que añaden características a los hechos donde la información es vista como cubos, los cuales consisten de categorías descriptivas (dimensiones) y valores cualitativos (medidas).

Realmente los cubos como tal no existen, los cubos se utilizan para representar o visualizar un modelo dimensional. Podemos representar un modelo dimensional con tres dimensiones con un cubo, en cualquier caso, lo habitual es tener modelos

con más de tres dimensiones, en ese caso su representación se realiza con cubos que reciben el nombre de hipercubos. Como los hipercubos son más difíciles de visualizar por eso se usa con mayor asiduidad el término cubo.

Operaciones OLAP. Recordemos que la granularidad de la información es directamente proporcional al nivel de detalle, por lo que estas operaciones básicamente nos ayudan a ver o analizar los datos desde diferentes niveles de granularidad. Las principales operaciones sobre modelos multidimensionales son:

- Drill-down: permite apreciar los datos con un mayor nivel de detalle (disgregación). Se aplica bajando por los niveles de una jerarquía definida en un cubo. Drill-down implica ir de lo general a lo específico.
- Roll-up: permite apreciar los datos con un menor nivel de detalle, subiendo por una jerarquía definida en un cubo. Esto brinda la posibilidad de quitar un nivel o criterio de agregación en el análisis. Roll-up implica ir de lo específico a lo general.
- Drill-across: es una operación que trabaja de forma similar a drill-down, con la diferencia de que drill-across NO se aplica sobre una jerarquía, sino que su forma de ir de lo general a lo específico es agregar un atributo a la consulta como nuevo criterio de análisis.
- Roll-across: es una operación que trabaja de forma similar a drill-up, con la diferencia de que roll-across NO se hace sobre una jerarquía, sino que su forma de ir de lo específico a lo general es quitar un atributo de la consulta, eliminando de esta manera un criterio de análisis.
- Slice & Dice: la operación de Slice realiza un “Corte” en el cubo para que los usuarios puedan centrarse en un área determinada del cubo, mientras que la operación Dice, rota el cubo hasta una nueva perspectiva, para que los usuarios puedan ver los datos desde diferentes aspectos en su análisis de los datos.
- Pivot: permite seleccionar el orden de visualización de los atributos e indicadores, con el objetivo de analizar la información desde diferentes puntos de vista.

Existen dos arquitecturas diferentes para los sistemas OLAP

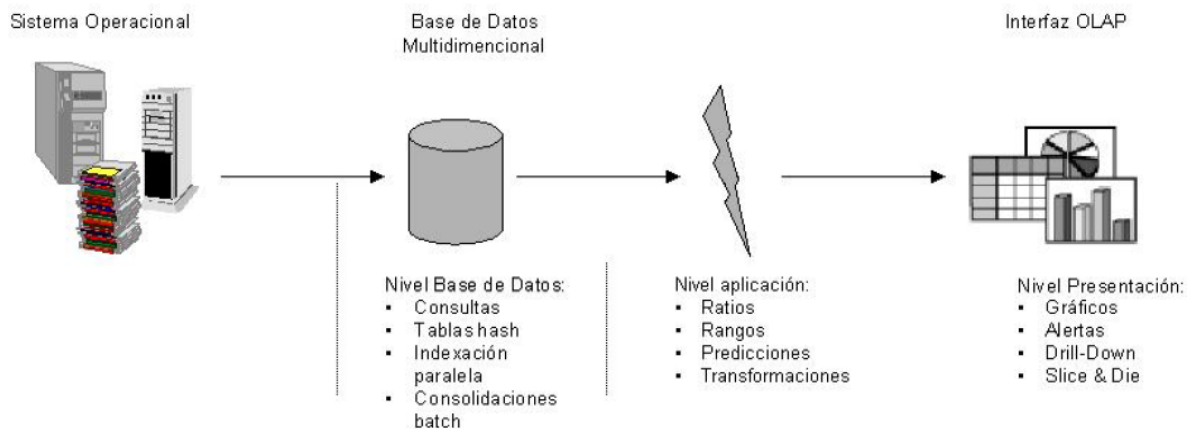
- MOLAP (Multidimensional Online Analytical Processing)
- ROLAP (Relational Online Analytical Processing)

Sistema MOLAP. En este sistema los datos se almacenan en forma lógica en arreglos matriciales con estructura multidimensional, además los datos son redundantes, por lo que se requiere mucha capacidad de almacenamiento y puede contener totales pre calculados (almacenados) es decir los resultados de todas las posibles consultas a la base de datos para optimizar los tiempos de respuesta.

MOLAP utiliza una arquitectura de dos niveles como se ve en la figura 3: la base de datos multidimensional y el motor analítico.

- La base de datos multidimensional es la encargada del manejo, acceso y obtención del dato.
- El nivel de aplicación es el responsable de la ejecución de los requerimientos OLAP. El nivel de presentación se integra con el de aplicación y proporciona una interfaz a través del cual los usuarios finales visualizan los análisis OLAP.

Figura 3. Representación sistema MOLAP (Multidimensional Online Analytical Processing).



Fuente:¹¹

¹¹ Ver Prototipo de mercado de datos para la división de control y fiscalización

En la tabla 1, miramos las ventajas y desventajas de utilizar un sistema MOLAP como técnica de explotación de un data warehouse.

Tabla 1. Ventajas y desventajas sistemas MOLAP

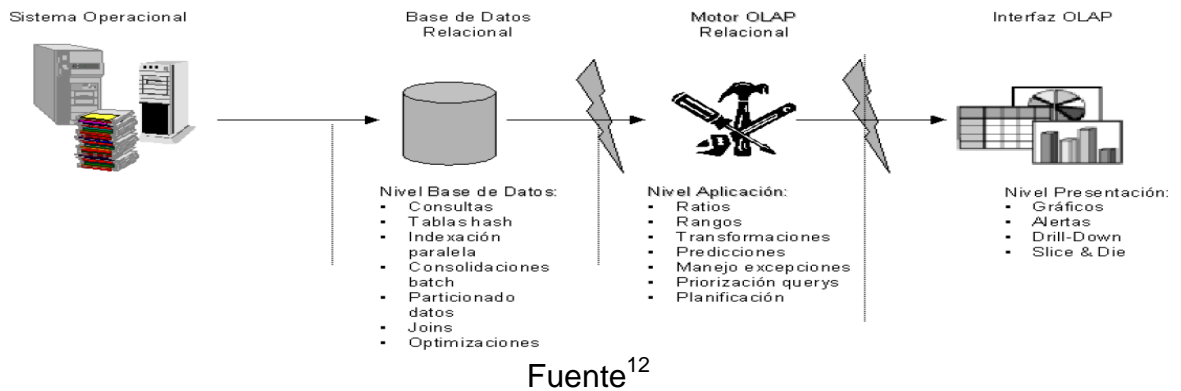
VENTAJAS	DESVENTAJAS
Mayor desempeño en el procesamiento de consultas.	Tamaño limitado para la arquitectura del cubo.
Se puede escribir sobre la base de datos.	No se permite acceder a los datos que están en el cubo.
Permite hacer cálculos más complejos.	No se permite explorar el paralelismo de la base de datos.

Fuente: Esta Investigación.

Sistema ROLAP. En este sistema los datos se encuentran almacenados en una base de datos relacional. Típicamente, los datos son detallados, evitando redundancia y las tablas se encuentran normalizadas. Los esquemas más comunes sobre los que se trabaja son en estrella o copo de nieve, aunque es posible trabajar sobre cualquier base de datos relacional.

En la figura 4 se observa el esquema de este sistema, donde el nivel de base de datos utiliza base de datos relacional para el manejo, acceso y obtención de los datos, el nivel de aplicación es el motor que ejecuta las consultas multidimensionales de los usuarios y el motor ROLAP se integra con niveles de presentación, a través de los cuales los usuarios realizan análisis OLAP.

Figura 4. Representación sistema ROLAP (Relational Online Analytical Processing).



En la tabla 2, miramos las ventajas y desventajas de utilizar un sistema ROLAP como técnica de explotación de un data warehouse.

Tabla 2. Ventajas y desventajas sistemas ROLAP.

VENTAJAS	DESVENTAJAS
Uso total de la seguridad e integridad de la base de datos.	Consultas más lentas.
Escalable para grandes volúmenes.	Construcción costosa.
Los datos pueden ser compartidos con aplicaciones de consulta SQL.	Los cálculos están limitados a las funciones de base de datos.
Datos y estructuras más dinámicos	

Fuente: Esta Investigación.

Desde el punto de vista del usuario, el único proceso visible es la explotación del data warehouse, aunque el éxito de construcción de un almacén de datos radica en los tres procesos iniciales que alimentan la información del mismo y suponen el mayor porcentaje de esfuerzo a la hora de desarrollar un data warehouse.

En la tabla 3, encontramos las diferencias que hay entre un sistema tradicional frente a un data warehouse.

Tabla 3. Diferencias de un datawarehouse con un sistema tradicional.

SISTEMA TRADICIONAL	DATA WAREHOUSE
Usuarios de perfiles medios o bajos	Usuarios de perfiles altos

¹² Ver Prototipo de mercado de datos para la división de control y fiscalización

Estructura relacional	Visión Multidimensional
Explotación de la información relacionada con la operativa de cada aplicación	Explotación de toda la información interna y externa relacionada con el negocio
Importancia del dato actual	Importancia del dato histórico
Importancia del tiempo de respuesta de la transacción instantánea	Importancia de la respuesta masiva
La actividad más importante es de tipo operativo (día a día)	La actividad más importante es el análisis y la decisión estratégica
Mayor importancia a la estabilidad	Mayor importancia al dinamismo
Datos en general desagregados	Datos en distintos niveles de detalle y agregación
Predomina la actualización	Predomina la consulta

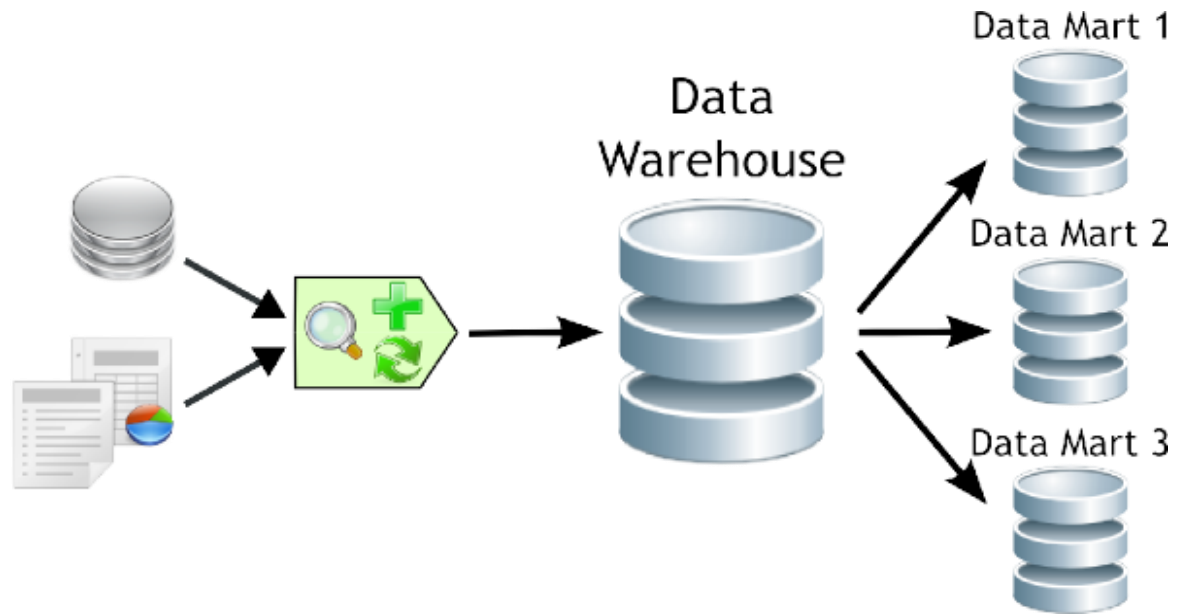
Fuente: Esta Investigación

Mercado de datos (Datamart)

Un mercado de datos es un subconjunto de un data warehouse, como se observa en la figura 5, es decir, es una base de datos departamental, especializada en consulta, análisis, almacenamiento e integración de los datos de un área de negocio específica cuya distribución interna de los datos es clara y no hay dudas al respecto, estando los datos estructurados en modelos multidimensionales de estrella o copo de nieve.

El mercado de datos se puede alimentar desde los datos de un data warehouse o de algunas fuentes distintas de información.

Figura 5. Esquema de un mercado de datos.



Fuente¹³

Los mercados de datos, son muy útiles para trabajar con herramientas OLAP para así ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones. Los datos existentes en este contexto pueden ser agrupados, explorados y propagados de múltiples formas para que diversos grupos de usuarios realicen la explotación de los mismos de la forma más conveniente según sus necesidades.

Data warehouse y mercado de datos. El objetivo de un data warehouse es extraer los datos de una o varias bases operacionales y moverlos a una base de datos independiente orientada a las consultas. Pero el problema surge cuando los data warehouse crecen y se tornan más complejos; en este caso la solución es crear la cantidad necesaria de mercados de datos, los cuales soportan menos usuarios que un data warehouse, por esta razón es posible recuperar más rápidamente los datos que necesitan los usuarios.

En la tabla 4 se hace una comparación de las características de un data warehouse frente a un mercado de datos.

Tabla 4. Data warehouse vs Mercado de datos.

¹³ Ver Riquelme, M. ¿Qué es un datamart? Web y empresas

DATA WAREHOUSE	MERCADO DE DATOS
Diseñado para optimizar la integración y la administración de los datos fuente.	Diseñado para optimizar la entrega de información de soporte a decisiones.
Administra grandes cantidades de datos históricos.	Se concentra en administrar resúmenes y/o datos totalizados.
Administrado por la unidad de sistemas de la organización.	Administrado por personal de sistemas de la unidad propietaria del mercado de datos.
Pertenece a toda la organización.	Pertenece al área de negocios al cual está orientado.
Construido para satisfacer las necesidades de información de toda la organización.	Construido para satisfacer las necesidades de información de toda la organización.

Fuente: Esta Investigación

Diseño de un mercado de datos. Existen múltiples metodologías de diseño y construcción de un mercado de datos, sin embargo, se imponen entre la mayoría dos de ellas, la de Kimball y la de Inmon donde la mayor diferencia entre los dos autores es el sentido de construcción del data warehouse.

Ralph Kimball presenta un enfoque Bottom-Up (de menor a mayor o ascendente), lo cual se ve reflejado en la implementación de un mercado de datos en áreas específicas de la organización, y poco a poco irlos integrando en un data warehouse corporativo, mientras que el enfoque de Bill Inmon es Top-Down (descendente), comenzando con todo el data warehouse desde el principio¹⁴

¿Qué metodología adoptar? La metodología a desarrollar en esta investigación es la de Kimball, por cuanto proporciona un enfoque de menor a mayor, muy versátil, y una serie de herramientas prácticas que ayudan a la implementación de un mercado de datos. El diseño se basa en la creación de tablas de hechos (FACTS), las cuales son tablas que contienen la información numérica de los indicadores a analizar, es decir la parte cuantitativa de la información.

¹⁴ Ver Rivadera, G. La metodología de Kimball para el diseño de almacenes de datos

Metodología Kimball

La metodología Kimball, se basa en lo que se denomina ciclo de vida dimensional del negocio (Business Dimensional Lifecycle), y nos indica que el data warehouse es un conglomerado de todos los mercados de datos dentro de una organización, con datos transaccionales, que son estructurados de una forma especial para el análisis.

Este enfoque se referencia como Bottom-up, pues al final el data warehouse corporativo no es más que la unión de los diferentes mercados de datos, que están estructurados de una forma común, lo cual permite que sea flexible y sencillo de implementar, pues existe la posibilidad de construir un mercado de datos como primer elemento del sistema de análisis, y luego ir añadiendo otros que comparten las dimensiones, ya definidas u otras nuevas¹⁵

La razón de ser de los proyectos de Business Intelligence y de muchos otros, es el negocio, por lo tanto, uno de los puntos importantes es tener claro que las necesidades del negocio son las que nos guiarán a lo largo de todo el proyecto. En general se contempla que el ciclo de vida dimensional del negocio propuesto por Kimball está basado en cuatro principios básicos:

- Centrarse en el negocio: Comprender con mucha claridad los requerimientos del negocio para poder traducirlos en un modelo de datos.
- Construir una infraestructura de información adecuada: Diseñar una base de información única, integrada, y fácil de manejar,
- Realizar entregas en incrementos significativos: Implementar en forma rápida y progresiva incrementos basados en procesos de negocio que conforman la matriz de datos empresariales.
- Ofrecer la solución completa: La entrega a los usuarios de una solución completa que incluya reportes, aplicaciones de consulta, portales, documentación, capacitación y soporte.

¹⁵ Ver Rivadera, G. La metodología de Kimball para el diseño de almacenes de datos

El ciclo de vida de Kimball como se observa en la figura 6, es una metodología detallada para el diseño, desarrollo e implementación de sistemas de BI y DW del cual podemos destacar dos cuestiones:

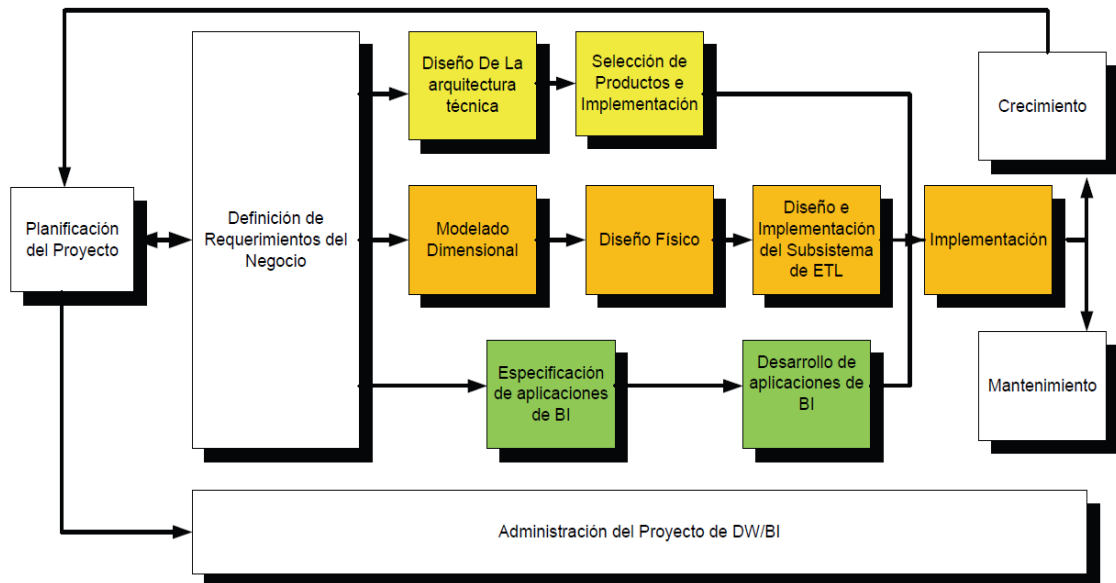
Primero hay que resaltar el rol central de la tarea de definición de requerimientos, los cuales son el soporte inicial de las tareas subsiguientes, y tiene influencia en el plan del proyecto.

En segundo lugar, podemos observar tres rutas o caminos que se enfocan en tres diferentes áreas.

- Tecnología: Implica tareas relacionadas con el software específico.
- Datos: Diseño e implementación del modelo dimensional y desarrollo de ETL.
- Aplicaciones de inteligencia de negocios: Diseño y desarrollo de aplicaciones de negocios para los usuarios finales.

Estas rutas se combinan cuando se instala finalmente el sistema.

Figura 6. Fases de la metodología Kimball.



Fuente¹⁶

¹⁶ Ver Kimball, R. & Ross, M. The data warehouse toolkit

Pruebas Saber 9°

Colombia tiene una amplia trayectoria en la realización de evaluaciones externas para conocer la calidad de la educación que están recibiendo los niños, niñas y jóvenes del país. Desde los primeros años de la década de los 90 se inició el desarrollo de las pruebas saber, concebido como una evaluación nacional de calidad educativa y se aplicaron sucesivamente pruebas a muestras de estudiantes de determinados grados y en algunas áreas del conocimiento, con el propósito de medir sus logros. Los resultados de estas primeras evaluaciones proporcionaron información valiosa para orientar la formulación de políticas y programas para apoyar el mejoramiento de la calidad educativa del país¹⁷

El Ministerio de Educación Nacional (MEN), con el propósito de garantizar que todos los estudiantes, independientemente de su procedencia y contexto socioeconómico y cultural en el que viven, reciban en la escuela una educación de alta calidad, que contribuya al desarrollo de las competencias necesarias para vivir, convivir, ser productivos en todos los ámbitos y seguir aprendiendo a lo largo de la vida¹⁸

Las pruebas saber, diseñadas por el Instituto para la Evaluación de la Educación (ICFES), e implementadas por la misma entidad, con el apoyo del MEN, son pruebas estandarizadas, a partir de la cual es posible evaluar lo que los estudiantes saben y saben hacer con lo que aprenden durante su trayectoria escolar y, a partir de sus resultados, identificar los avances, así como las deficiencias o debilidades sobre las cuales es preciso tomar medidas para mejorar. Con ello, el seguimiento escolar puede concretarse en metas claras y factibles, lo que a su vez fortalece la gestión del sector al hacer operativo su seguimiento y la evaluación del impacto de las acciones implementadas¹⁹

Los resultados en las pruebas saber, son un insumo para la construcción del índice sintético de calidad, que mide, junto con otras variables como el progreso de las instituciones, la eficiencia, el ambiente escolar y el desempeño de los colegios, así como para definir e implementar programas y acciones de mejoramiento en los establecimientos educativos, las secretarías y el Ministerio de Educación Nacional.

¹⁷ Ver Informe técnico Saber 5° y 9°

¹⁸ Ver Alineación del examen Saber 11° Lineamientos generales 2014-2

¹⁹ Ver Alineación del examen saber 11°. Sistema nacional de evaluación estandarizada

Esta evaluación, además, permite hacer un seguimiento desde que la escolaridad está empezando, para que la evaluación sea formativa. Se trata de que no ocurra que en el último año se reciben malos resultados, que impiden el acceso a la universidad, pero ya no hay nada que hacer, porque la escolaridad terminó²⁰

¿Cuál es el objetivo de Saber 9°? El objetivo de las pruebas Saber es contribuir al mejoramiento de la calidad de la educación colombiana, mediante la realización de evaluaciones censales periódicas en la 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, el MEN y la sociedad en general conozcan cuáles son las fortalezas y debilidades y, a partir de las mismas, 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 y establecer el impacto de programas y acciones específicas de mejoramiento.

¿A quiénes se evalúa? Saber, evalúa la calidad de la educación en la totalidad de los establecimientos educativos, tanto oficiales como privados, mediante la aplicación, de pruebas de competencias básicas a los estudiantes de 9° grado, así como de cuestionarios que recogen información sobre los contextos personales, familiares y escolares, a través de los cuales se busca conocer los factores que explican los resultados obtenidos.

¿Qué se evalúa? Las pruebas valoran las competencias que han desarrollado los estudiantes de 8° y 9° grado, acorde con los estándares básicos de competencias establecidos por el Ministerio de Educación Nacional.

Las características de las pruebas no abarcan la totalidad de contenidos ni de estándares definidos para cada área, no permiten evaluar la totalidad de las competencias que se espera que desarrollen los estudiantes, pero sus resultados son indicadores fundamentales del avance en el proceso de formación en educación básica y permiten establecer los niveles de desempeño, las fortalezas y las debilidades de los estudiantes para, a su vez, plantear estrategias concretas de mejoramiento.

²⁰ Ver Saber 5° y 9° 2009 Síntesis de resultados de factores asociados .

Estructura y alineación de las pruebas saber 9°. El ICFES para el año 2009 diseñó las pruebas saber 3°, 5° y 9° de tal forma que garanticen evaluaciones censales para un periodo de doce años con el fin de visualizar en los resultados la evolución que ha tenido la educación en Colombia. Estas pruebas evalúan las competencias en matemáticas, lenguaje y ciencias naturales²¹

El diseño de las pruebas fue realizado por docentes expertos en el área y se tuvo en cuenta los estándares básicos de competencias, tomando como un referente común en cuanto a lo que saben y saben hacer los estudiantes durante su proceso educativo, independientemente de su lugar de origen, modalidad educativa, condiciones culturales, sociales, económicas, religiosas, entre otras.

En la tabla 5, se evidencia la conformación de las pruebas saber 2009 teniendo en cuenta las competencias en lenguaje, matemáticas y ciencias naturales.

Tabla 5. Procesos de las competencias genéricas.

LENGUAJE	MATEMÁTICAS	CIENCIAS
<ul style="list-style-type: none"> • Lectura. • Escritura. 	<ul style="list-style-type: none"> • Razonamiento y argumentación. • Comunicación, representación y modelación. • Planteamiento y resolución de problemas. 	<ul style="list-style-type: none"> • Uso comprensivo del conocimiento científico. • Explicación de fenómenos. • Indagación.

Fuente: Guía de lineamientos generales pruebas saber 2009

Además de esto, en cada una de estas áreas se evalúan sus componentes, los cuales están relacionados con los ejes verticales de los estándares básicos de competencias como se muestra en la tabla 6. Esto permitirá establecer las fortalezas y debilidades de los estudiantes y, a partir de ellas, plantear estrategias concretas de mejoramiento.

Tabla 6. Componentes de las competencias genéricas.

LENGUAJE	MATEMÁTICAS	CIENCIAS
----------	-------------	----------

²¹ Ver Informe técnico Saber 5° y 9° 2009. Instituto Colombiano para la Evaluación de la Educación

Semántica	Numérico - variacional	Entorno vivo
Sintaxis	Geométrico - métrico	Entorno físico
Pragmática	Aleatorio	Ciencia, tecnología y sociedad (CTS)

Fuente: Guía de lineamientos generales pruebas saber 2009

La prueba comprende preguntas de selección múltiple con única respuesta, las cuales presentan un enunciado con cuatro opciones de respuesta A, B, C, D. Solo una de ellas es la correcta y valida con respecto a la situación planteada. El número de preguntas que contesta cada estudiante y en particular los de grado noveno son: 54 para lenguaje, 54 en matemáticas y 54 para ciencias naturales²²

Ahora bien, teniendo en cuenta el Plan Nacional Decenal de Educación 2006-2016, el ICFES ha avanzado en la alineación del Sistema Nacional de Evaluación Externa Estandarizada (SNEE), a través de la reestructuración de los exámenes: en 2009 con un nuevo diseño de saber 3°, 5° y 9°; en 2010 con el rediseño de saber Pro; en 2014 con los cambios en saber 11°. La alineación posibilita la comparación de los resultados en distintos niveles educativos, ya que diferentes exámenes evalúan unas mismas competencias en algunas áreas que los conforman, a saber, las competencias genéricas.²³

En el caso particular de las pruebas saber 9°, que se aplica desde el segundo semestre del 2014, la alineación consistió en la introducción de una prueba de competencias ciudadanas la cual presenta situaciones de análisis que se relacionan con su entorno educativo, es decir, el aula, el colegio, la familia y el barrio²⁴

La Ex Provincia de Obando

La ex provincia de Obando está ubicada en la zona sur del departamento de Nariño, en límites con el Ecuador. Está conformada por trece municipios; Aldana, Cuaspud, Guachucal, Cumbal, Gualmatan, Iles, Contadero, Ipiales, Pupiales, Funes, Córdoba, Puerres y Potosí. Posee una extensión de 4.894 kilómetros cuadrados aproximadamente, que equivalen al 14.07% del área total del departamento. Su población es de 296.026 habitantes que corresponden al

²² Ver Marco de factores Asociados Saber 3°, 5° y 9° 2016

²³ Ver Lineamientos generales Saber 2009 5° y 9°

²⁴ Ver Saber 5° y 9° Síntesis de resultados de factores asociados

16.49% del total del departamento; de los cuales 143.611 están ubicados en el sector urbano y 152.415 en el sector rural.

Aspectos demográficos. Según el último censo del Departamento Administrativo Nacional de Estadística (DANE) en 2005, en la tabla 7 se muestra las proyecciones de la población por cada municipio para el año 2017.

Tabla 7. Población municipios de la Ex Provincia de Obando.

Municipio	Urbano / Habitantes	Rural / Habitantes	Total, Población / Habitantes
Aldana	1.278	4.651	5.929
Contadero	2.386	4.617	7.003
Córdoba	2.375	11.691	14.066
Cuaspud	2.405	6.284	8.689
Cumbal	8.792	30.274	39.066
Funes	1.977	4.425	6.402
Guachucal	2.973	12.437	15.410
Gualmatan	2.245	3.533	5.778
Iles	1.982	6.889	8.871
Ipiales	105.875	39.198	145.073
Potosí	2.152	9.780	11.932
Puerres	3.066	5.193	8.259
Pupiales	6.105	13.443	19.548

Fuente²⁵

²⁵ Ver Proyección de Población municipales por área

METODOLOGÍA

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

Planificación del proyecto

La planificación busca identificar la definición y el alcance del proyecto de DW/BI, sus objetivos específicos, evaluaciones de factibilidad, los principales riesgos y una aproximación inicial a las necesidades de la información.

Según sentencia Kimball *“Antes de comenzar un proyecto de data warehouse o mercado de datos, hay que estar seguro si existe la demanda y de donde proviene. Si no se tiene un sólido usuario sponsor y no hay usuarios entusiasmados, posponga el proyecto”*.²⁶

Esta fase incluye las siguientes acciones típicas de un plan de proyecto:

- Definir el alcance (entender los requerimientos del negocio).
- Identificar las tareas.
- Programar las tareas.
- Planificar el uso de los recursos.
- Asignar la carga de trabajo a los recursos.
- Elaboración de un documento final que representa un plan del proyecto.

Análisis de requerimientos

Un factor determinante en el éxito de un proceso de data warehouse es la interpretación correcta de los diferentes niveles de requerimientos expresados por los diferentes niveles de usuarios.

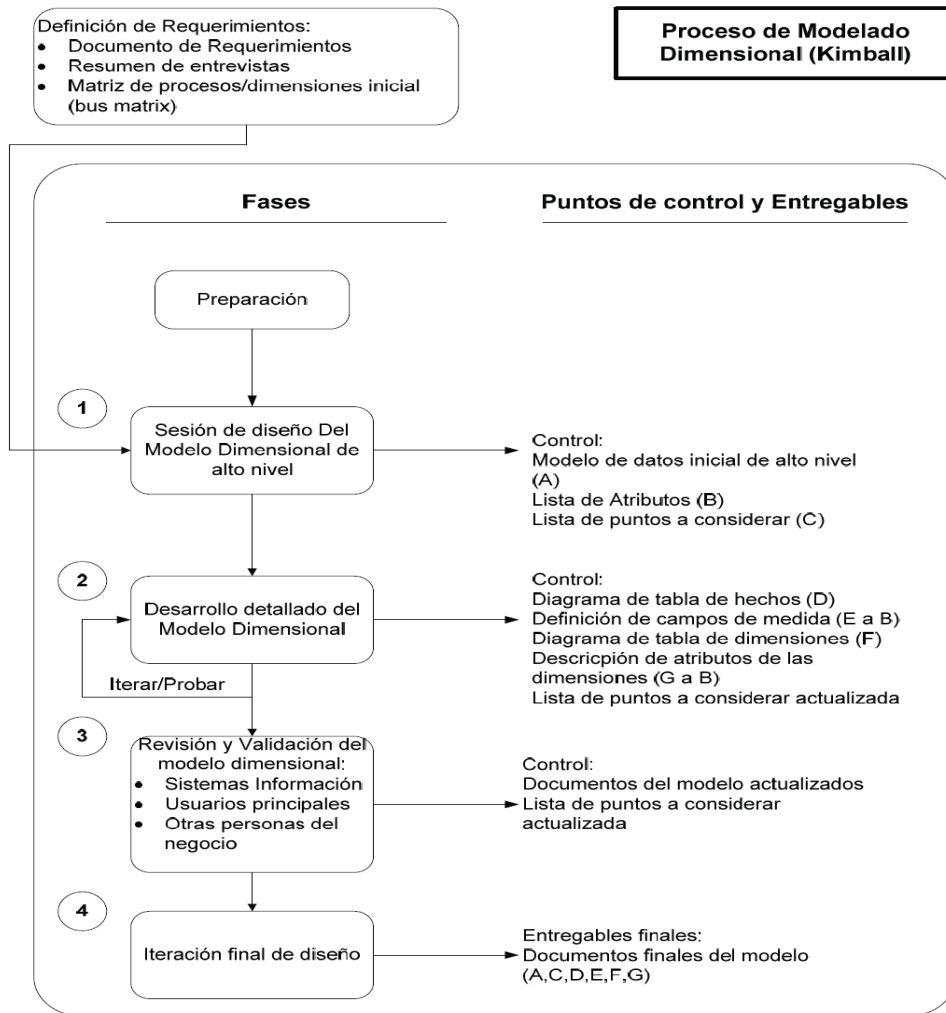
Según la perspectiva de Kimball, los requerimientos del negocio se posicionan en el centro del *“universo del data warehouse”* y determinan el alcance del mismo (que datos debe contener, como debe estar organizado, cada cuanto debe actualizarse, quienes y desde donde accederán, etc.)

²⁶ Ver Mundy & Thornthwaite. The Microsoft data warehouse Toolkit

Modelado dimensional

El proceso de diseño empieza con un modelo dimensional de alto nivel obtenido a partir de los procesos priorizados de la matriz de requerimientos. En la figura 7 se muestra el diagrama de flujo del proceso de modelado dimensional de Kimball, el cual es un proceso dinámico y altamente iterativo.

Figura 7. Diagrama de flujo del proceso dimensional de Kimball.



Fuente²⁷

El proceso iterativo consiste en cuatro pasos:

²⁷ Ver Kimball, R, & Ross, M, The Data Warehouse Toolkit

1. Elegir el proceso de negocio: El primer paso es elegir el área a modelar, depende fundamentalmente del análisis de requerimientos.
2. Establecer el nivel de granularidad: La granularidad significa especificar el nivel de detalle, depende de los requerimientos del negocio y lo que es posible a partir de los datos actuales. Diseñar el mercado de datos a mayor nivel de detalle permite realizar agrupamientos al nivel deseado
3. Elegir las dimensiones: Las dimensiones surgen naturalmente de las discusiones del equipo y del nivel de granularidad. Las tablas de dimensiones tienen un conjunto de atributos (generalmente textuales) que brindan una perspectiva o forma de análisis sobre una medida en una tabla de hechos. Estos atributos son posibles candidatos para ser encabezado en los informes, tablas, cubos, o cualquier forma de visualización, unidimensional o multidimensional.
4. Identificar medidas y las tablas de hechos: El último paso consiste en identificar las medidas que surgen de los procesos de negocio. Una medida es un atributo (campo) de una tabla que se desea analizar y se encuentran en tablas que se denominan tablas de hechos (FACT) y cada tabla tiene como atributos una o más medidas, de acuerdo a los requerimientos.

Diseño Físico

Se focaliza sobre la selección de las estructuras necesarias para soportar el diseño lógico. Algunos de los elementos principales de este proceso son la definición de convenciones estándares de nombre específicos del ambiente de la base de datos.

Diseño del subsistema de Extracción, Transformación y Carga (ETL)

Estas tareas son altamente críticas, pues tienen que ver con la materia prima del data warehouse: los datos. Es por ello que la calidad de los datos es un factor determinante en el éxito de un proyecto de data warehouse.

Como advierte Kimball, el proceso ETL es el iceberg de un proyecto de data warehouse. Son muchos los desafíos que deben enfrentarse para lograr datos de

alta calidad de los sistemas fuentes. En general, es una de las etapas más subestimadas que siempre termina tomando más tiempo de lo previsto.

Especificación y desarrollo de aplicaciones de BI

Una parte fundamental de todo proyecto DW/BI está en proporcionar a una gran comunidad de usuarios, una forma más estructurada de la información y por lo tanto más fácil de acceder al almacén de datos, esto es posible mediante las aplicaciones de inteligencia de negocios.

Los reportes y aplicaciones de BI son la cara visible de la inteligencia de negocios, proporcionan información útil a los usuarios e incluyen un amplio espectro de tipos de reportes y herramientas de análisis que van desde reportes simples de formato fijo a sofisticadas aplicaciones analíticas que utilizan algoritmos complejos.

Kimball divide a estas aplicaciones en dos categorías basadas en el nivel de sofisticación, y les llama reportes estándar y aplicaciones analíticas.

Reportes Estándar. Son reportes relativamente simples, de formato predefinido y parámetros de consulta fijos, es decir reportes estáticos almacenados. Estos reportes estándar proporcionan a los usuarios un conjunto básico de información acerca de lo que está sucediendo en cierta área de la organización.

Aplicaciones analíticas. Estas aplicaciones son más complejas que los reportes estándar puesto que se centran en un proceso de negocio específico y resumen cierta experiencia de cómo analizar e interpretarlo. Estas aplicaciones pueden ser muy avanzadas e incluir algoritmos y modelos de minería de datos, que ayudan a identificar oportunidades o cuestiones subyacentes en los datos.

1. IMPLEMENTACIÓN

Para el desarrollo del proyecto se da inicio al Ciclo de Vida Dimensional del Negocio (Business Dimensional Lifecycle) comenzado con el análisis de requerimientos.

1.1. Análisis de Requerimientos

El interés de este proyecto se centró en construir un mercado de datos de los resultados obtenidos por los estudiantes de las instituciones educativas de los municipios de la subregión de Obando, en las pruebas saber 9°, y de esta forma contar con una herramienta de inteligencia de negocios que permita desarrollar análisis multidimensional que genere beneficios a los directivos, y secretarías de educación de estos municipios.

- Disponer de un repositorio de los resultados de las pruebas saber 9° para el periodo comprendido entre los años 2014 a 2016.
- Analizar el desempeño histórico de los estudiantes en las pruebas saber 9° para el periodo comprendido entre los años 2014 a 2016.

Para darle solución al porcentaje de valores nulos, se dividió los datos en tablas según las competencias que se hayan aplicado en determinado año en las pruebas saber.

La tabla 8 contiene 16 filas cada una de ellas corresponde a las posibles combinaciones de las competencias genéricas que se pueden dar en su aplicación. Se obtuvieron un total de 16 tablas, de las cuales se omitieron siete debido a que su combinación no se podía dar.

Tabla 8. Consolidado de las posibles combinaciones entre competencias genéricas

N° de fila	Lenguaje	Matemáticas	Ciencias	Competencias	Numero de datos
1	0	0	0	0	0
2	0	0	0	1	55
3	0	0	1	0	55
4	0	0	1	1	0
5	0	1	0	0	156
6	0	1	0	1	922
7	0	1	1	0	2154

8	0	1	1	1	0
9	1	0	0	0	226
10	1	0	0	1	980
11	1	0	1	0	2195
12	1	0	1	1	0
13	1	1	0	0	3136
14	1	1	0	1	0
15	1	1	1	0	0
16	1	1	1	1	0

Fuente: Esta investigación.

La fila 1 corresponde a la tabla que muestra que no se aplicó ninguna competencia, lo que no es posible, porque en los años estudiados deben aplicarse tres de las cuatro competencias genéricas.

Las filas 2, 3, 5 y 9 se relacionan con las tablas que aparecen con una sola competencia aplicada y corresponden a competencias ciudadanas, ciencias naturales, matemáticas, lenguaje respectivamente. Estos datos se obviaron en el proceso puesto que no cumple con las condiciones iniciales planteadas en los lineamientos de las pruebas.

La fila 4 se obvio puesto que la combinación de las competencias ciudadanas y de ciencias naturales no se aplican simultáneamente si no en años separados y sumado a esto n contiene datos.

La fila 8 muestra la aplicación de las competencias: matemáticas, competencias ciudadanas y ciencias naturales, lo cual no es posible por lo mencionado en el párrafo anterior y por esto no contiene valores.

La fila 12 muestra la aplicación de tres competencias: lenguaje, ciudadanas y ciencias naturales, lo cual no es posible por lo mencionado anteriormente y por esto no contiene valores.

La fila 14 y 15 se obvio porque representa la combinación de tres competencias y esto no se puede dar.

La fila 16 se excluye puesto que no se puede aplicar las cuatro competencias al mismo tiempo.

Finalmente, solo quedaron las filas 6, 7, 10, 11 y 13 que corresponden a la aplicación de competencias ciudadanas y matemáticas, ciencias naturales y matemáticas, competencias ciudadanas y lenguaje, ciencias naturales y lenguaje,

y matemáticas y lenguaje, respectivamente; Con cada una de estas filas se construyó una tabla que conservan los atributos de la tabla de estudiantes.

1.2. Modelado dimensional

Para la construcción del mercado de datos se contó inicialmente con datos de 2.365.142 estudiantes elegidos de manera aleatoria de las diferentes instituciones académicas y un total de 72 atributos, que en términos generales pueden ser agrupados en las siguientes categorías:

- Características generales del estudiante.
- Características de los municipios.
- Características de los establecimientos educativos.
- Características del tipo prueba.
- Características del ente territorial.
- Características del tiempo.
- Características de las sedes.

En la tabla 9 se muestra las dimensiones con su nivel de granularidad que se utilizó para construir el modelado multidimensional.

Tabla 9. Dimensiones y granularidad del mercado de datos.

Dimensión	Granularidad
Estudiante	Estudiante
Municipios	Municipios (subregión de Obando)
Establecimientos	Establecimientos educativos
Ente Territorial	Nombre del ente territorial
Tiempo	Año que se realizó la prueba
Sedes	Sedes de los establecimientos educativos
Tipo Prueba	Prueba (evaluada por áreas del conocimiento)

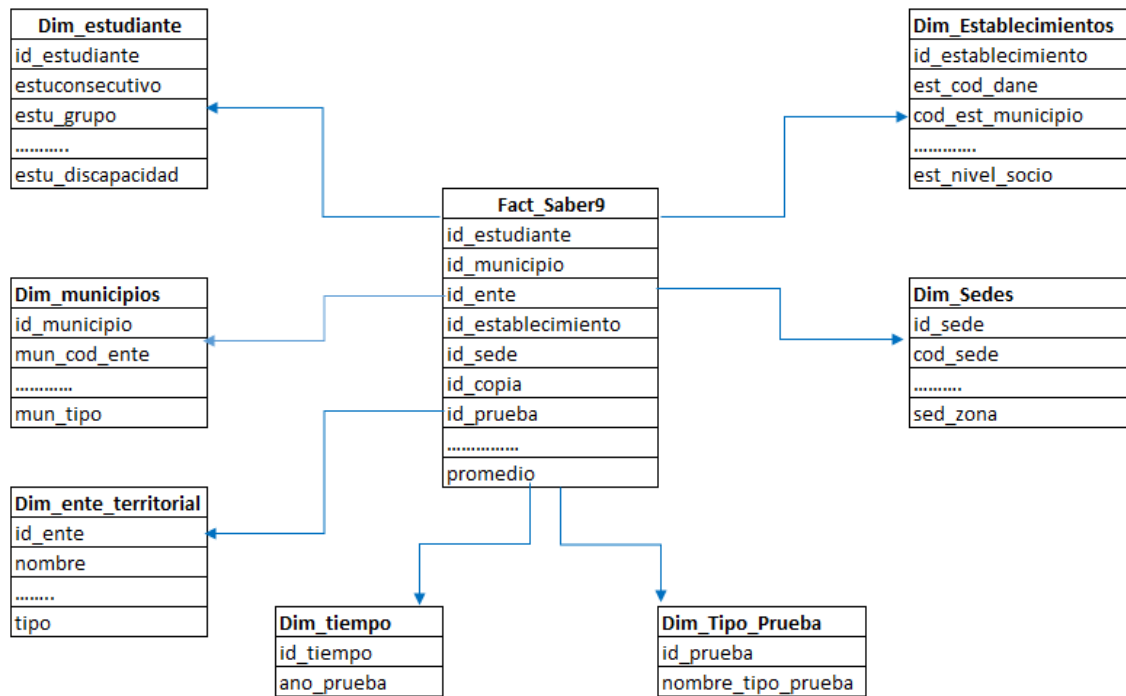
Fuente: Esta investigación.

Se tiene una tabla de hechos (Fact_Saber9) que tiene como granulo el estudiante, y se identificaron las siguientes medidas sobre los hechos:

- ¿Cuántos estudiantes presentaron las pruebas saber 9° en el periodo 2014 a 2016?
- ¿Cuántos estudiantes presentaron las pruebas de las ciencias básicas del saber (lenguaje-ciencias, lenguaje-competencias, lenguaje-matemáticas, matemáticas-ciencias y matemáticas-competencias) en el periodo 2014 a 2016?
- ¿Cuál es el promedio del puntaje obtenido en las ciencias básicas del saber (lenguaje-ciencias, lenguaje-competencias, lenguaje-matemáticas, matemáticas-ciencias y matemáticas-competencias) en el periodo 2014 a 2016?

En la figura 8 se muestra un esquema en estrella, el cual es un tipo de base de datos relacional que se compone de una única tabla de hechos central, rodeada por tablas de dimensiones, siendo este un esquema para el diseño del modelo multidimensional asociado al mercado de datos.

Figura 8. Esquema en estrella.



Fuente: Esta investigación

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 relacionales PostgreSQL versión 9.5.1; inicialmente se usaron tres repositorios de datos suministrados por el FTP del ICFES, cada uno con información referente a las pruebas saber 9° para los años 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. El propósito general de esta fase fue construir un repositorio de datos limpio y estandarizado a partir de las tres tablas fuentes entregadas; para tal fin se desarrollaron tres guías de trabajo.

Integración de datos. Para esta fase se siguieron los siguientes procedimientos:

1. Descargar de la página ICFES los repositorios de las pruebas saber 9°, los cuales contienen los resultados obtenidos por los estudiantes de las instituciones educativas de la subregión de Obando en los años 2014, 2015 y 2016.
2. Se creó la base de datos Saber9 en el SGBD PostgreSQL.
3. En la base de datos Saber9 se creó las siguientes tablas:
 - Plausibles (2014, 2015, 2016).
Esta tabla almacena la identificación del estudiante, año de aplicación de la prueba, estrato al que pertenece el establecimiento, grado del estudiante, etc. Y los atributos que la componen se muestran en la tabla 10
 - comp_estab (2014, 2015,2016).
Esta tabla almacena el nombre del establecimiento educativo, la zona donde se ubica el establecimiento, urbano o rural, etc. Los atributos que la componen se muestran en la tabla 11
 - comp_ent_terr (2014, 2015,2016).

Esta tabla almacena las entidades territoriales certificadas y los atributos que la componen se muestra en la tabla 12

- comp_municipio (2014, 2015,2016).
Esta tabla contiene el nombre de los municipios y los atributos que la componen se muestra en la tabla 13
 - comp_sede (2014, 2015,2016).
Esta tabla contiene código saber de la sede jornada, código DANE de la sede del establecimiento educativo y los atributos que la componen se muestra en la tabla 14.
4. Se realiza la carga de los archivos planos al gestor SGBD PostgreSQL según las tablas anteriores. Ver anexo 1

Tabla 10. Valores Plausibles.

VARIABLES	2014	2015	2016	DESCRIPCIÓN DE VARIABLES
estu_consecutivo	X	X	X	Identificación del estudiante asignado por el ICFES.
aplicación	X	X	X	Año de aplicación de la prueba
calendario	X	X	X	Calendario del establecimiento
departamento	X	X	X	Código DANE del departamento al que pertenece el establecimiento educativo
dissenso	X	X	X	Marca de discapacidad cognitiva
enteterr	X	X	X	Código DANE de la entidad a la que pertenece el establecimiento educativo
establecimiento	X	X	X	Código DANE del establecimiento educativo al que pertenece el estudiante
Estrato	X	X	X	Estrato al que pertenecen el establecimiento dentro del marco muestral
Grado	X	X	X	Grado del estudiante
Grupo	X	X	X	Grupo/salón al que pertenecen los estudiantes
id_hoja	X	X	X	Código de identificación de hoja de respuestas
jornada	X	X	X	Código de la sede jornada al que pertenece el estudiante
leng_copietas	X	X	X	Indicador de copia en lenguaje
leng_score1	X	X	X	Score1-valor plausible 1 de lenguaje

Tabla 10 (Continuación)

VARIABLES	2014	2015	2016	DESCRIPCIÓN DE VARIABLES
leng_score3	X	X	X	Score3-valor plausible 3 de lenguaje
leng_score4	X	X	X	Score4-valor plausible 4 de lenguaje
leng_score5	X	X	X	Score5-valor plausible 5 de lenguaje
leng_weight	X	X	X	Peso muestral del lenguaje
mate_copietas	X	X	X	Indicador de copia en matemáticas
mate_score1	X	X	X	Score1-valor plausible 1 de matemáticas
mate_score2	X	X	X	Score2-valor plausible 2 de matemáticas
mate_score3	X	X	X	Score3-valor plausible 3 de matemáticas
mate_score4	X	X	X	Score4-valor plausible 4 de matemáticas
mate_score5	X	X	X	Score5-valor plausible 5 de matemáticas
mate_weight	X	X	X	Peso muestral de matemáticas
cien_copietas	X			Indicador de copia en ciencias
cien_score1	X			Score1-valor plausible 1 de ciencias
cien_score2	X			Score2-valor plausible 2 de ciencias
cien_score3	X			Score3-valor plausible 3 de ciencias
cien_score4	X			Score4-valor plausible 4 de ciencias
cien_score5	X			Score5-valor plausible 5 de ciencias
cien_weight	X			Peso muestral de ciencias
comp_copietas		X	X	Indicador de copia en competencias
comp_score1		X	X	Score1-valor plausible 1 de competencias
comp_score2		X	X	Score2-valor plausible 2 de competencias
comp_score3		X	X	Score3-valor plausible 3 de competencias
comp_score4		X	X	Score4-valor plausible 4 de competencias
comp_score5		X	X	Score5-valor plausible 5 de competencias
comp_weight		X	X	Peso muestral de competencias
modeloedu	X	X	X	No está definido en el diccionario de datos
municipio	X	X	X	Código DANE al que pertenece el establecimiento educativo
N	X	X	X	Matricula del grupo al que pertenecen
Nivel	X	X	X	Nivel socioeconómico del establecimiento
sector	X	X	X	Sector del establecimiento
Sexo	X	X	X	Sexo del estudiante
Zona	X	X	X	Zona de la jornada
Zonastab	X	X	X	Zona del establecimiento

Fuente: Esta investigación.

Tabla 11. Establecimientos.

VARIABLES	2014	2015	2016	DESCRIPCIÓN DE VARIABLES
cod_dane	X	X	X	Código DANE del establecimiento educativo
id_municipio	X	X	X	Código DANE del municipio al que pertenece el establecimiento educativo
id_ente	X	X	X	id de la entidad territorial
Nombre	X	X	X	Nombre del establecimiento educativo reportado en el DUE
Zona	X	X	X	Zona donde la mayoría de la población atendida por el establecimiento educativo se ubica 1=urbano 2=rural
Sector	X	X	X	Naturaleza administrativa de establecimiento educativo según 1=oficial urbano 2=oficial rural 3=no oficial
tipo_estab	X	X	X	Tipo de establecimiento 1=Oficial urbano 2=Oficial rural 3=no oficial
Calendario	X	X	X	Calendario del establecimiento
nivel_socio	X	X	X	NSE asignado de acuerdo a la clasificación realizada con puntajes por medios

Fuente: Esta investigación.

Tabla 12. Entidades.

VARIABLES	2014	2015	2016	DESCRIPCIÓN DE VARIABLES
id_ente	X	X	X	Identificador del ente territorial
Nombre	X	X	X	Nombre del ente territorial
Munexclu	X	X	X	Nombre de los municipios certificados excluidos dentro del ente territorial
Tipo	X	X	X	Tipo de ente territorial 1=ETC, 2=DPTO,3=MPIO

Fuente: Esta investigación.

Tabla 13. Municipio.

VARIABLES	2014	2015	2016	DESCRIPCIÓN DE VARIABLES
id_municipio	X	X	X	Código DANE del municipio según divipo
Nombre	X	X	X	Nombre del municipio según divipo

Fuente: Esta investigación.

Tabla 14. Sedes.

VARIABLES	2014	2015	2016	DESCRIPCIÓN DE VARIABLES
Id	X	X	X	Código saber de la sede jornada
codigo_dane_estab	X	X	X	Código DANE del establecimiento educativo
codigo_dane_sede	X	X	X	Código DANE de la sede del establecimiento educativo
Jornada	X	X	X	Jornada en la que desempeña actividades y para la que se reporta resultados. M=Mañana, T=Tarde, C=Completa
Nombre	X	X	X	Nombre de la sede jornada según DUE

Fuente: Esta investigación.

Análisis de la calidad de datos

Se crearon las tablas rep_plau_inicial2014_2016, rep_enteterr2014_2016, rep_municipio2014_2016, rep_sedes2014_2016, rep_establecimiento2014_2016 a partir de la tabla comp_municipio, comp_estab, comp_ent_terr, comp_sede (2014, 2015, 2016) con el propósito de realizar un análisis inicial de la calidad de los datos e identificar el número de registros no nulos, nulos, distintos valores y valores distintos, máximos y mínimos.

En la tabla 15 se muestran los resultados del análisis de calidad de datos de valores plausibles.

Tabla 15. Análisis de la calidad de valores plausibles.

ATRIBUTO	# NO NULOS	#NULOS	#VALORES DISTINTOS	VALORES DISTINTOS	MAX	MIN
estu_consecutivo	1686063	0	1686063	"V201490000001" "V201490000002" "V201490000003" "V201490000004" "V201490000005" "V201490000006" "V201490000007" "V201490000008" "V201490000009" "V201490000010"	V201690571717	V201490000001
Grupo	1686063	0	12	"1" "10" "11" "2" "3" "4" "5" "6" "7" "8"	1	99
n	1686063	0	381	"1" "10" "100" "101" "102" "103" "104" "105" "106" "107"	99	1
estrato	1686063	0	28	"00000MA" "02E9999" "02O9909" "02O9919" "11" "12E9999" "12O1909" "12O2909" "21" "391909"	99O2999	00000MA
aplicación	1686063	0	1	X	x	x
grado	1686063	0	1	9	9	9

Tabla 15 (Continuación)

jornada	1686063	0	13592	"178230" "178263" "178290" "178293" "178308" "178332" "178335" "178422" "178428" "178587"	999999	178230
establecimiento	1686063	0	13059	"105001000000" "105001000001" "105001000043" "105001000132" "105001000141" "105001000167" "105001000175" "105001000205" "105001000256" "105001000353"	8,54874E+11	1,05001E+11
enteterr	1686063	0	95	"1" "10" "11" "12" "13" "14" "15" "16" "17" "18"	95	1
departamento	1686063	0	33	"11" "13" "15" "17" "18" "19" "20" "23" "25" "27"	99	11
municipio	1686063	0	1118	"11001" "13001" "13006" "13030" "13042" "13052" "13062" "13074" "13140" "13160"		

Tabla 15 (Continuación)

sexo	1686063	0	3	"1" "2" "3"	1	3
sector	1686063	0	2	"1" "2"	2	1
zona	1686063	0	2	"1" "2"	2	1
zonastab	1686063	0	2	"1" "2"	2	1
calendario	1686063	0	2	"A" "B"	B	B
nivel	1686063	0	5	"1" "2" "3" "4" "5"	5	1
modeloedu	1686063	0	1	X	x	x
dissenso	1686063	0	2	"0" "1"	1	0
leng_copietas	1115874	570189	3	"0" "1" "2" ""	2	0
leng_weight	1115874	570189	2370	"1" "1,05" "1,05556" "1,05882" "1,06" "1,07" "1,08" "1,1" "1,10345" "1,14"	6	1
mate_copietas	1110446	575617	4	"0" "1" "2" "3" ""	3	0
mate_weight	1110446	575617	2370	"1" "1,04762" "1,05556" "1,06" "1,08" "1,1" "1,13333" "1,14" "1,14286" "1,16667"	6	1

Tabla 15 (Continuación)

cien_copietas	739352	946711	4	"0" "1" "2" "3" ""	3	0
cien_weight	739352	946711	1664	"1" "1,14" "1,15" "1,17" "1,2" "1,22" "1,23" "1,25" "1,27" "1,28571"	6	1
comp_copietas	356908	1329155	2	"0" "1" ""	1	0
comp_weight	356908	1329155	1587	"1" "1,05" "1,2" "1,25" "1,26316" "1,26471" "1,26667" "1,28571" "1,3" "1,30556"	6	1
ano	1686063	0	3	"2014" "2015" "2016"	2014	2016

Fuente: Esta investigación.

En la tabla 16 se muestra el análisis de calidad de datos de establecimientos.

Tabla 16. Análisis de la calidad de establecimientos.

ATRIBUTO	# NO NULOS	# NULOS	#VALORES DISTINTOS	VALORES DISTINTOS	MAX	MIN
cod_dane	23717	0	23717	"105001000001" "105001000043" "105001000108" "105001000132" "105001000141" "105001000167" "105001000175" "105001000205" "105001000256" "105001000353"	925754004540	105001000001
id_municipio	23717	0	1121	"11001" "13001" "13006" "13030" "13042" "13052" "13062" "13074" "13140" "13160"	99773	11001
id_ente	23717	0	95	"1" "10" "11" "12" "13" "14" "15" "16" "17" "18"	95	1
nombre	23717	0	20935	"ACAD LA SALLE SAN BENILDO" "ACAD LEGION BRITANICA" "ACAD MILIT MARISCAL SUCRE" "ACAD MILIT NAL HEROES GRANADINOS" "ACAD PANTANO DE VARGAS" "ACAD. MILITAR GRAL. CUSTODIO GARCIA R." "ACADE MILIT JOAQUÍN DE CAICEDO Y CUERO" "ACADEMIA SAM" "ACADEMIA ARTISTICA CHAMALU" "ACADEMIA CIAL DE LOS ANDES"	ZULDEMAYDA	ACAD LA SALLE SAN BENILDO

Tabla 16 (Continuación)

zona	23717	0	2	"1" "2"	2	1
sector	23717	0	2	"1" "2"	2	1
tipo_estab	23717	0	3	"1" "2" "3"	3	1
calendario	23717	0	2	"A" "B"	B	A
nivel_socio	23717	0	5	"1" "2" "3" "4" "5"	9	1

Fuente: Esta investigación.

En la tabla 17 se muestra el análisis de calidad de datos del ente territorial.

Tabla 17. Análisis de la calidad de Ente Territorial.

ATRIBUTO	# NO NULOS	# NULOS	#VALORES DISTINTOS	VALORES DISTINTOS	MAX	MIN
id_ente	99	0	97	"1" "10" "11" "12" "13" "14" "15" "16" "17" "18"	97	1
nombre	99	0	98	"ALICIA " "ALICIA EN EL PAIS" "AMAZONAS" "ANTIOQUIA" "APARTADO" "ARAUCA" "ARMENIA" "ATLANTICO" "BARRANCABERMEJA" "BARRANQUILLA"	ZIPAQUIRA	ALICIA
muneclu	98	0	26	"ARMENIA" "BUCARAMANGA, BARRANCABERMEJA, FLORIDABLANCA, GIRON,	YOPAL	ARMENIA

				PIEDECUESTA" "CALI, BUENAVENTURA, BUGA, CARTAGO, PALMIRA, TULUA, JAMUNDI" "CALI, BUENAVENTURA, BUGA, CARTAGO, PALMIRA, TULUA, JAMUNDI, YUMBO" "CIENAGA, SANTA MARTA" "CUCUTA" "FLORENCIA" "FUSAGASUGA, GIRARDOT, SOACHA, CHIA, FACATATIVA, ZIPAQUIRA, MOSQUERA" "IBAGUE" "MAGANGUE, CARTAGENA"		
tipo	99	0	3	"ARMENIA" "BUCARAMANGA, BARRANCABERMEJA, FLORIDABLANCA, GIRON, PIEDECUESTA" "CALI, BUENAVENTURA, BUGA, CARTAGO, PALMIRA, TULUA, JAMUNDI" "CALI, BUENAVENTURA, BUGA, CARTAGO, PALMIRA, TULUA, JAMUNDI, YUMBO" "CIENAGA, SANTA MARTA" "CUCUTA" "FLORENCIA" "FUSAGASUGA, GIRARDOT, SOACHA, CHIA, FACATATIVA, ZIPAQUIRA, MOSQUERA" "IBAGUE" "MAGANGUE, CARTAGENA"	2	0

Fuente: Esta investigación.

En la tabla 18 se muestran los resultados del análisis de calidad de datos de municipios.

Tabla 18. Análisis de la calidad de Municipio.

ATRIBUTO	# NO NULOS	# NULOS	#VALORES DISTINTOS	VALORES DISTINTOS	MAX	MIN
id_ente	1061	0	1061	"13006" "13030" "13042" "13052" "13062" "13074" "13140" "13160" "13188" "13212"	99773	13006
nombre	1061	0	978	"ABEJORRAL" "ABREGO" "ABRIAQUI" "ACACIAS" "ACANDI" "ACEVEDO" "ACHI" "AGRADO" "AGUA DE DIOS" "AGUACHICA"	ZONA BANANERA	ABEJORRAL
tipo	1061	0	1	4	4	4

Fuente: Esta investigación.

En la tabla 19 se muestra los resultados del análisis de calidad de datos de sedes

Tabla 19. Análisis de la calidad de Sedes.

ATRIBUTO	# NO NULOS	# NULOS	#VALORES DISTINTOS	VALORES DISTINTOS	MAX	MIN
id	84424	0	84421	"178116" "178188" "178230" "178263" "178290" "178293" "178305" "178308" "178332" "178335"	674356	178116

codigo_dane_estab	84424	0	17153	"105001000001" "105001000043" "105001000108" "105001000132" "105001000141" "105001000167" "105001000175" "105001000205" "105001000256" "105001000353"	854874000001	105001000001
codigo_dane_sede	84424	0	54875	"102050000022" "105001000001" "105001000027" "105001000043" "105001000078" "105001000108" "105001000132" "105001000141" "105001000159" "105001000167"	927413000001	102050000022
jornada	84424	0	3	"C" "M" "T"	T	C
nombre	84424	0	42576	"# 1" "# 1 CLEMENCIA" "# 1 DE CORDOBA" "# 1 DE SANTA CRUZ" "# 1 SAN JUAN" "# 2" "# 2 BARRIO SUR" "# 2 CALDERA" "# 2 CLEMENCIA" "# 2 DE GUAMO - SAN ANTONIO"	ZULUAGUITA	# 1

Fuente: Esta investigación.

Limpieza de datos

Se ejecutaron varios procesos sobre las tablas comp_ent_terr2014_2016, comp_estab2014_2016, comp_municipio2014_2016 y plausibles2014_2016, con el propósito de estandarizar los valores presentes, rellenar los valores nulos, corregir valores erróneos y en general mejorar las condiciones de calidad del repositorio. Ver Anexo 1.

1.3.2. Fase de Transformación. El objetivo de esta fase fue construir un repositorio estandarizado y con datos de calidad que fue usado como fuente de carga del mercado de datos; se desarrollaron dos guías de trabajo. Ver Anexo 2.

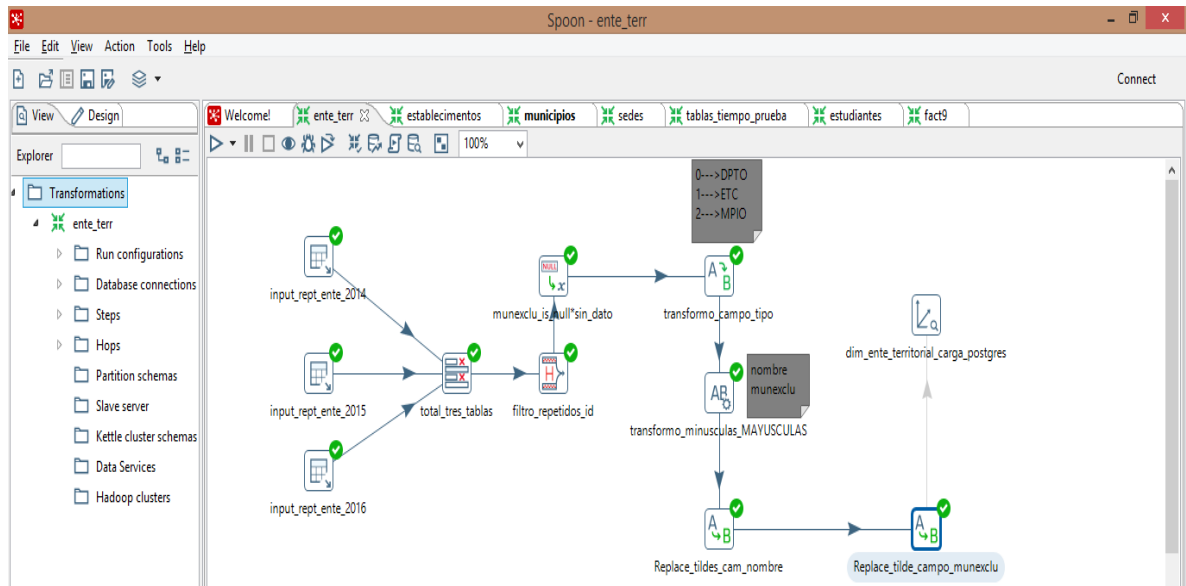
- **Transformación de datos.** A partir de los repositorios rep_plau_inicial2014_2016, rep_enteterr2014_2016, rep_municipio2014_2016, rep_sedes2014_2016, rep_establecimiento2014_2016 se crearon las tablas rep_plau2014_2016_final, rep_enteterr2014_2016_final, rep_municipio2014_2016_final, rep_sedes2014_2016_final, rep_establecimiento2014_2016_final, sobre las cuales se ejecutaron procedimientos para normalizar el valor de los atributos y facilitar el posterior análisis de resultados.
- **Eliminación de datos.** A partir de los repositorios rep_plau_inicial2014_2016, rep_enteterr2014_2016, rep_municipio2014_2016, rep_sedes2014_2016, rep_establecimiento2014_2016 se crearon las tablas rep_plau2014_2016_final, rep_enteterr2014_2016_final, rep_municipio2014_2016_final, rep_sedes2014_2016_final, rep_establecimiento2014_2016_final, sobre las cuales se efectuaron rutinas para eliminar atributos que no son requeridos en el mercado de datos.

1.3.3. Fase de Carga. Para el proceso de carga en el mercado de datos, solo se consideraron aquellos registros relevantes para el fin de esta investigación; a partir de la bodega de datos saber_9 se creó la base de datos final llamado data_mart, el cual contiene las dimensiones y tablas de hechos. Posteriormente se procede a integrar los datos con la herramienta PDI (Pentaho Data Integration) a las 7 tablas del mercado de datos (7 dimensiones y 1 tabla de hechos).

A continuación, se presenta el proceso ETL automatizado utilizando la herramienta Pentaho Data Integration.

- Proceso ETL automatizado para la dimensión ente territorial el cual se muestra en la figura 9.

Figura 9. Proceso ETL automatizado en la herramienta Pentaho Data Integration (PDI) el cual fue utilizado en la carga de datos de la dimensión ente territorial.



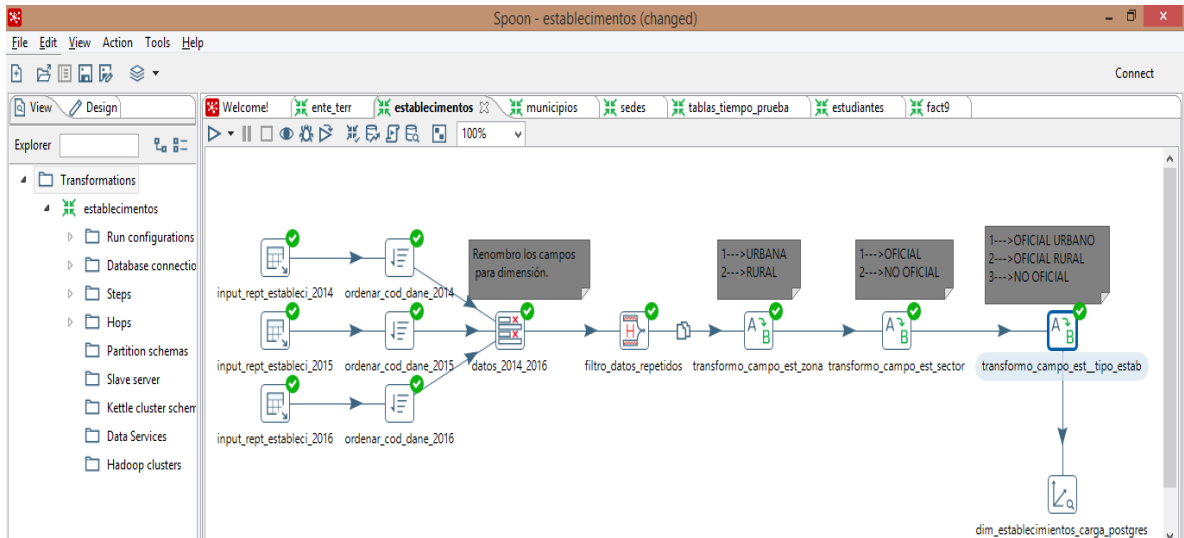
Fuente: Esta investigación.

En la figura 9 se observa el proceso ETL de la dimensión ente territorial:

- Inicialmente se cargó los repositorios comp_ent_terr(2014,2015,2016) a los nodos 'Table input' correspondiente a cada año .
- En el nodo 'select values' se combinan los datos de las tres tablas.
- En el nodo 'Unique' row se toman los datos únicos a través del campo id_ente.
- En el nodo 'value is null' en el campo munexclu se reemplazó los valores nulos por SIN_DATO.
- En el nodo 'Replace in string' en el campo tipo se reemplazó los valores 1=DPTO, 2=ETC, 3=MPIO.
- En el nodo 'String operations' en los campos nombre, munexclu reemplazo los valore 1=DPTO, 2=ETC, 3=MPIO.
- En el nodo 'Replace in string' en los campos nombre, munexclu se estandarizo todo tipo de tildes.
- Finalmente, en el nodo 'Dimension lookup/update' se realiza la carga a la dimensión dim_ente_territorial en el gestor postgres en la base de datos data_mart.

- Proceso ETL automatizado para la dimensión establecimientos.

Figura 10. Proceso ETL automatizado en la herramienta Pentaho Data Integration (PDI) el cual fue utilizado en la carga de datos de la dimensión establecimientos.



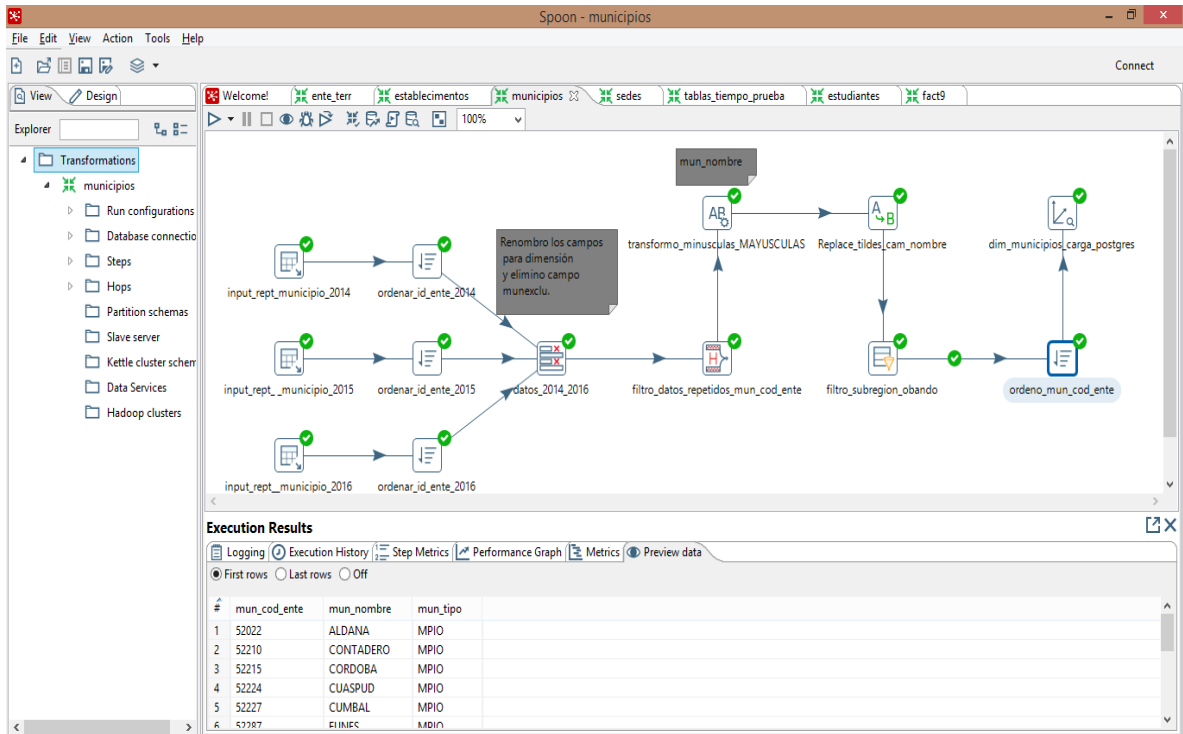
Fuente: Esta investigación.

En la figura 10 se observa el proceso ETL de la dimensión establecimientos:

- Inicialmente se cargó los repositorios comp_estab(2014,2015,2016) a los nodos 'Table input' correspondiente a cada año.
- En los nodos 'Sort rows' por cada entrada se ordenó los datos ascendentemente a través del campo cod_dane.
- En el nodo 'select values' se combinan los datos de las tres tablas y renombro los campos originales por los nombres mencionados en la dimensión dim_establecimientos.
- En el nodo 'Unique row' se toman los datos únicos a través del campo est_cod_dane.
- En el nodo 'Replace in string' en el campo est_zona se reemplazó los valores 1=URBANA, 2=RURAL.
- En el nodo 'Replace in string' en el campo est_sector se reemplazó los valores 1=OFICIAL, 2=NO OFICIAL.
- Finalmente, en el nodo 'Dimension lookup/update' se realiza la carga a la dimensión dim_establecimientos en el gestor postgres en la base de datos data_mart.

- Proceso ETL automatizado para la dimensión municipios.

Figura 11. Proceso ETL automatizado en la herramienta Pentaho Data Integration (PDI) el cual fue utilizado en la carga de datos de la dimensión municipios.



Fuente: Esta investigación.

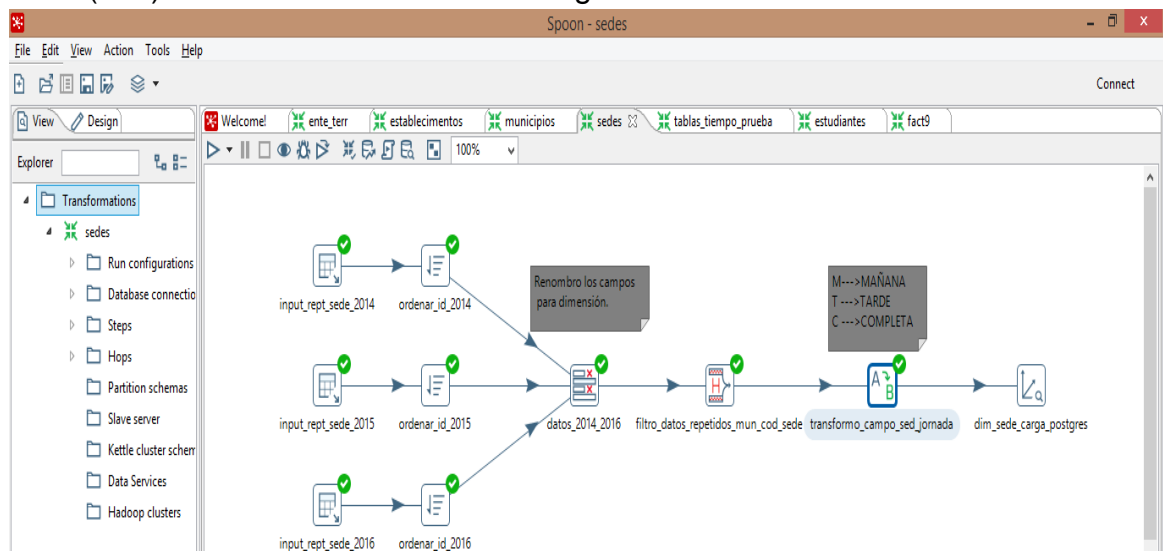
En la figura 11 se observa el proceso ETL de la dimensión municipios:

- Inicialmente se cargó los repositorios comp_municipio(2014,2015,2016) a los nodos 'Table input' correspondiente a cada año.
- En los nodos 'Sort rows' por cada entrada se ordenó los datos ascendentemente a través del campo id_ente.
- En el nodo 'select values' se combinan los datos de las tres tablas y renombro los campos originales por los nombres mencionados en la dimensión dim_municipios.
- En el nodo 'Unique' row se toman los datos únicos a través del campo mun_cod_ente.
- En el nodo 'Replace in string' en el campo mun_nombre se reemplazó letras minúsculas a mayúsculas para tener mejor estandarización.
- En el nodo 'Replace in string' en los campos mun_nombre se estandarizo todo tipo de tildes.

- En el nodo 'Filter rows' se filtran los datos correspondientes a los municipios de la subregión de Obando a través del campo mun_cod_ente de acuerdo al código que le corresponde a cada municipio.
- Finalmente, en el nodo 'Dimension lookup/update' se realiza la carga a la dimensión dim_municipios en el gestor postgres en la base de datos data_mart.

- Proceso ETL automatizado para la dimensión sedes.

Figura 12. Proceso ETL automatizado en la herramienta Pentaho Data Integration (PDI) el cual fue utilizado en la carga de datos de la dimensión sedes.



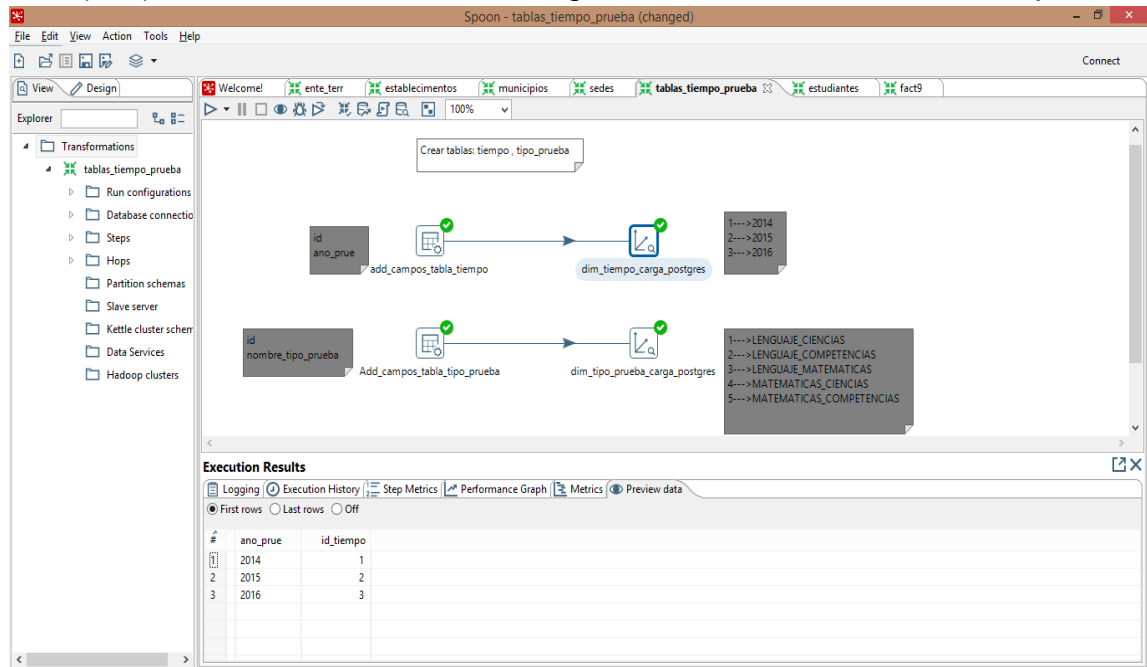
Fuente: Esta investigación.

En la figura 12 se observa el proceso ETL de la dimensión sedes:

- Inicialmente se cargó los repositorios comp_sede(2014,2015,2016) a los nodos 'Table input' correspondiente a cada año.
- En los nodos 'Sort rows' por cada entrada se ordenó los datos ascendentemente a través del campo **id**.
- En el nodo 'select values' se combinan los datos de las tres tablas y renombro los campos originales por los nombres mencionados en la dimensión dim_sedes.
- En el nodo 'Unique row' se toman los datos únicos a través del campo sed_codigo_dane_sede.
- En el nodo 'Replace in string' en el campo **sed_jornada** se reemplazó los valores M=MANANA, T=TARDE, C=COMPLETA.

- Finalmente, en el nodo ‘Dimension lookup/update’ se realiza la carga a la dimensión dim_sedes en el gestor postgres en la base de datos data_mart.
- Proceso ETL automatizado para la dimensión tiempo.

Figura 13. Proceso ETL automatizado en la herramienta Pentaho Data Integration (PDI) el cual fue utilizado en la carga de datos de la dimensión tiempo.

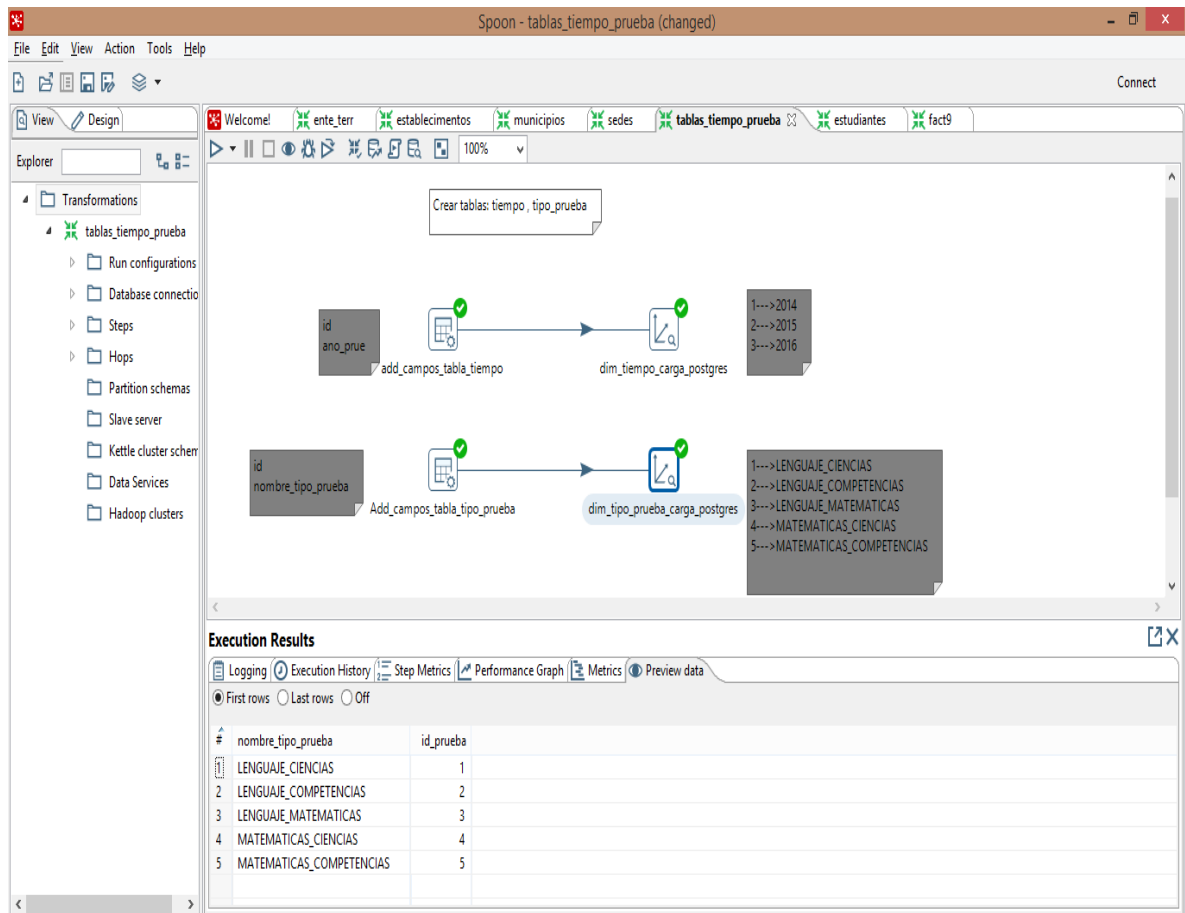


Fuente: Esta investigación.

En la figura 13 se observa el proceso ETL de la dimensión tiempo:

- Se tomó el nodo ‘Data Grid’ para la creación de la tabla tiempo con sus respectivos campos **id_tiempo** y **ano_prue** con los siguientes valores 1=2014, 2=2015 y 3=2016.
- Finalmente, en el nodo ‘Dimension lookup/update’ se realiza la carga a la dimensión dim_tiempo en el gestor postgres en la base de datos data_mart.
- Proceso ETL automatizado para la dimensión tipo prueba.

Figura 14. Proceso ETL automatizado en la herramienta Pentaho Data Integration (PDI) el cual fue utilizado en la carga de datos de la dimensión tipo prueba.

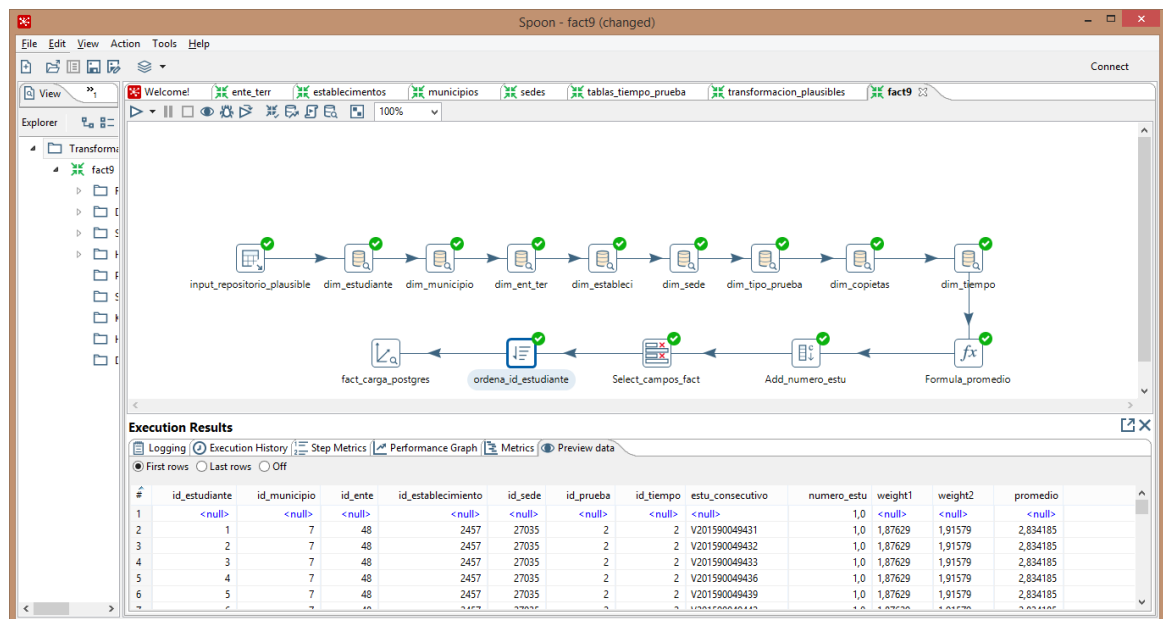


Fuente: Esta investigación.

En la figura 14 se observa el proceso ETL de la dimensión tipo de prueba:

- Se tomó el nodo 'Data Grid' para la creación de la tabla tipo prueba con sus respectivos campos **id_prueba** y **nombre_tipo_prueba** con los siguientes valores 1=LENGUAJE_CIENCIAS, 2=LENGUAJE_COMPETENCIAS, 3=LENGUAJE_MATEMATICAS, 4=MATEMATICAS_CIENCIAS Y 5=MATEMATICAS_COMPETENCIAS.
- Finalmente, en el nodo 'Dimension lookup/update' se realiza la carga a la dimensión dim_tipo_prueba en el gestor postgres en la base de datos data_mart.
- Proceso ETL automatizado para la tabla de hechos.

Figura 15. Proceso ETL automatizado en la herramienta Pentaho Data Integration (PDI) el cual fue utilizado en la carga de datos de la tabla de hechos.



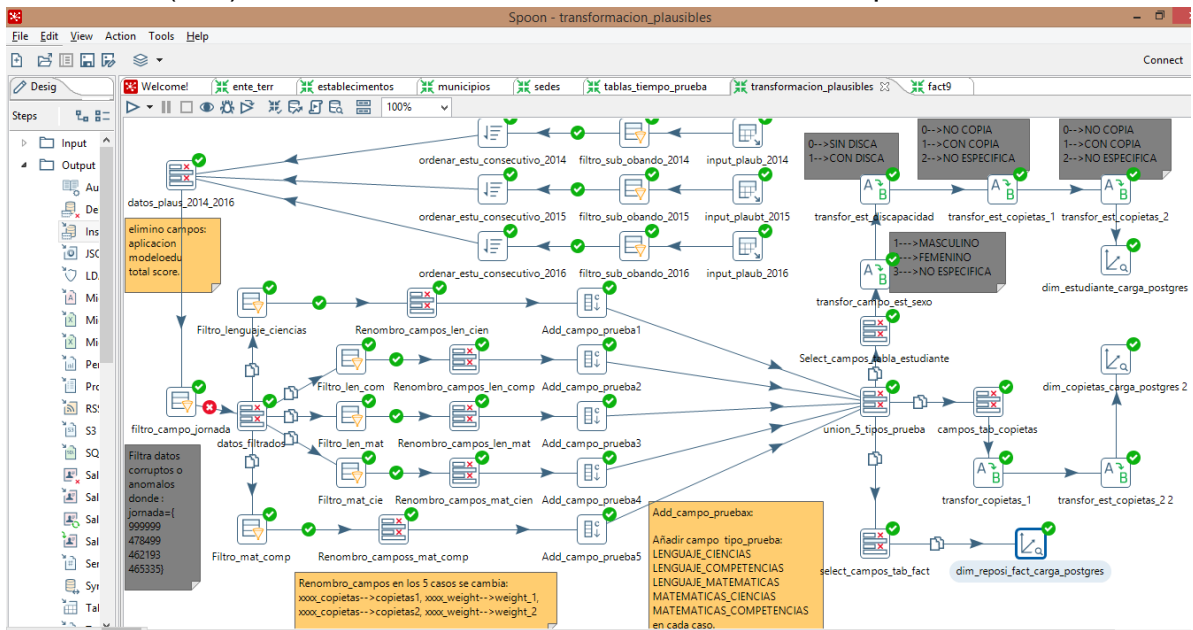
Fuente: Esta investigación.

En la figura 15 se observa el proceso ETL de la tabla de hechos:

- Inicialmente se cargó los repositorios Plausibles (2014,2015,2016) a los nodos 'Table input' correspondiente a cada año.
- En los nodos 'Filter rows' por cada año se filtran los datos correspondientes a los municipios de la subregión de Obando a través del campo municipio de acuerdo al código correspondiente.
- En los nodos 'Sort rows' por cada entrada se ordenó los datos ascendentemente a través del campo estu_consecutivo.
- En el nodo 'select values' se combinan los datos de las tres tablas y elimino los campos aplicación, modeloedu y scores.
- En el nodo 'Filter rows' se filtran los datos anómalos a través del campo jornada con los códigos iguales a: 999999, 478499, 462193, 465335.
- En el nodo 'select values' selecciono los datos filtrados.
- En los nodos 'Filter rows' se filtran los datos de acuerdo al tipo de prueba correspondientes a: LENGUAJE_CIENCIAS, LENGUAJE_COMPETENCIAS, LENGUAJE_MATEMATICAS, MATEMATICAS_CIENCIAS, MATEMATICAS_COMPETENCIAS través de los campos copietas y weight.
- En los nodos 'select values' se renombra los campos originales copietas por los nombres mencionados en la dimensión dim_copietas.
- En los nodos 'Add constants' se incluye el campo tipo_prueba y el valor con el nombre que le corresponde de acuerdo a cada tipo prueba.
- En el nodo 'select values' se combinan los datos de las cinco tablas clasificadas por su tipo de prueba.

- Por un lado, a mono derecha en el nodo 'select values' selecciono los campos estu_consecutivo, copietas1, copietas2 para finalmente en el nodo 'Dimension lookup/update' realizar la carga a la dimensión dim_tipo_prueba en el gestor postgres en la base de datos data_mart.
 - Por otro lado, el nodo 'select values' que apunta hacia arriba es utilizado para seleccionar los campos que tendrá la dimensión dim_estudiante.
 - En el nodo 'String operations' en el campo estu_sexo se remplazó los valores 1=MASCULINO, 2=FEMENINO, 3=NO_ESPECIFICA.
 - En el nodo 'String operations' en el campo estu_discapacidad se remplazó los valores 1= SIN_DISCAPACIDAD, 2=CON_DISCAPACIDAD.
 - Finalmente, en el nodo 'Dimension lookup/update' se realiza la carga a la dimensión dim_estudiante en el gestor postgres en la base de datos data_mart.
- Proceso ETL automatizado para la transformación de plausibles

Figura 16. Proceso ETL automatizado en la herramienta Pentaho Data Integration (PDI) el cual fue utilizado en la transformación de plausibles.



Fuente: Esta investigación.

En la figura 16 se observa el proceso ETL de la tabla de hechos:

- Inicialmente se cargó repositorio_plausibles al nodo 'Table input'.
- En los nodos 'Database lookup' se cargó cada dimensión para integrar la tabla de hechos con su respectivo id.
- En el nodo 'formula' se calcula el promedio de weight1 y weight2.

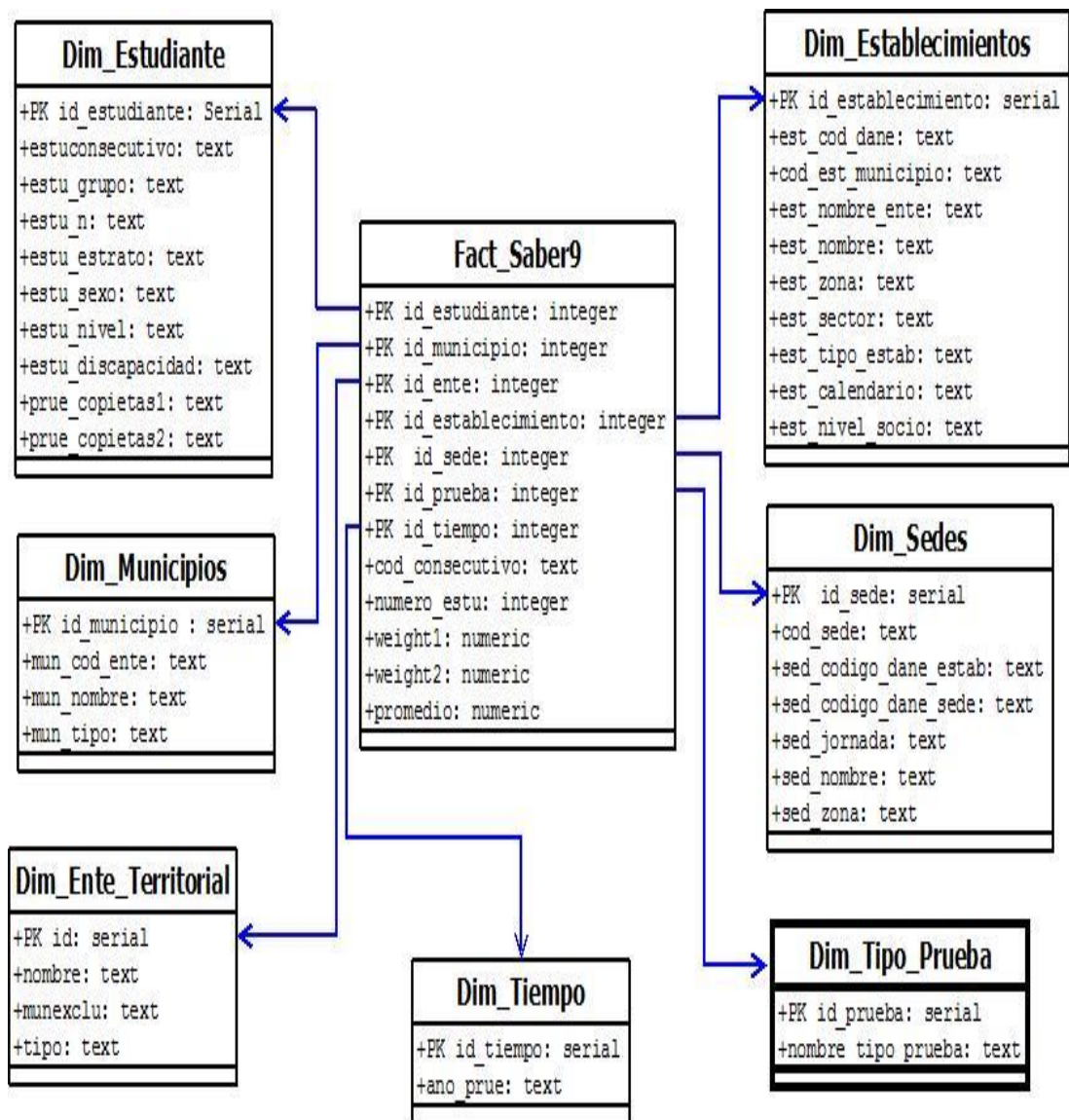
- En el nodo 'Add constants' se incluye el campo numero_estu con el valor igual a 1.
- En el nodo 'Sort rows' se ordenó los datos ascendentemente a través del campo id_estudiante.
- Finalmente, en el nodo 'Dimension lookup/update' se realiza la carga a la tabla de hechos fact_saber9 en el gestor postgres en la base de datos data_mart.

1.4. Implementación del mercado de datos en PostgreSQL

Para la implementación del mercado de datos Saber 9^o cuyo modelo dimensional en estrella que se muestra en la figura 8, se utilizó el sistema gestor de base de datos PostgreSQL. Tanto las dimensiones como los hechos se convirtieron en tablas relacionales en un modelo ROLAP.

En la figura 17 se muestra la implementación del modelo multidimensional tipo estrella en PostgreSQL, que cuenta con 8 dimensiones con sus respectivas llaves primarias y las variables relevantes correspondientes a cada una, además este modelo está constituido por una tabla de hechos con su respectiva llave primaria, sus llaves foráneas y las variables relevantes.

Figura 17. Implementación del Modelo Multidimensional tipo estrella en PostgreSQL.



Fuente: Esta investigación.

1.5. Diccionario del mercado de datos

Tabla 20. Diccionario de datos de la dimensión Estudiante.

No.	ATRIBUTO	DESCRIPCIÓN	TIPO
PK	id_estudiante	Llave primaria	serial
1	estuconsecutivo	Identificador del estudiante	text
2	estu_grupo	Salón al que pertenece el estudiante	text
3	estu_n	Número de matriculados en el grupo al que pertenecen	text
4	estu_estrato	Estrato al que pertenece el establecimiento dentro del marco muestral	text
5	estu_sexo	Género estudiante	text
6	estu_nivel	Nivel Socioeconómico (NSE) asignado de acuerdo a la clasificación realizada con puntajes promedios: NSE 1, NSE 2, NSE 3, NSE 4, 5 Sin NSE	text
7	estu_discapacidad	Con o Sin discapacidad	text
8	prue_copietas1	Indicador de copia de matemáticas: No copia, Con copia	text
9	prue_copietas2	Indicador de copia de matemáticas: No copia, Con copia	text

Fuente: Esta investigación.

Tabla 21. Diccionario de datos de la dimensión Sedes.

No.	ATRIBUTO	DESCRIPCIÓN	TIPO
PK	id_sede	Llave primaria	serial
1	cod_sede	Código Saber de la sede-jornada	text
2	sed_codigo_dane_estab	Código DANE del Establecimiento Educativo	text
3	sed_codigo_dane_sede	Código DANE de la sede del Establecimiento Educativo	text
4	sed_jornada	Jornada en la que desempeña actividades: Mañana, Tarde, Completa	text
5	sed_nombre	Nombre de la sede según el DUE	text
6	sed_zona	Zona donde pertenece la sede	text

Fuente: Esta investigación.

Tabla 22. Diccionario de datos de la dimensión Establecimientos.

No.	ATRIBUTO	DESCRIPCIÓN	TIPO
PK	id_establecimiento	Llave primaria	serial
1	est_cod_dane	Código DANE del Establecimiento Educativo	text
2	id_est_municipio	Código DANE del municipio al que pertenece el Establecimiento Educativo	text
3	est_nombre_ente	Nombre del ente territorial	text
4	est_nombre	Nombre del Establecimiento Educativo reportado en el DUE	text
5	est_zona	Zona donde el Establecimiento Educativo se ubica: Urbano, Rural	text
6	est_sector	Naturaleza administrativa del Establecimiento Educativo según el DUE: Oficial, No Oficial	text
7	est_tipo_estab	Tipo de establecimiento. Oficial urbano, Oficial rural, No Oficial	text
8	est_calendario	Calendario del Establecimiento: Calendario A, Calendario B	text
9	est_nivel_socio	Nivel Socioeconómico asignado de acuerdo (NSE): NSE 1, NSE 2, NSE 3, NSE 4, 5 Sin NSE	text

Fuente: Esta investigación.

Tabla 23. Diccionario de datos de la dimensión Tipo Pruebas.

No.	ATRIBUTO	DESCRIPCIÓN	TIPO
PK	id_prueba	Llave primaria	serial
1	nombre_tipo_prueba	Nombre de las dos pruebas que realizó el estudiante.	text

Fuente: Esta investigación.

Tabla 24. Diccionario de datos de la dimensión Ente territorial.

No.	ATRIBUTO	DESCRIPCIÓN	TIPO
PK	id	Llave primaria	serial
1	Nombre	Nombre del ente territorial	text
2	munexclu	Nombre de los municipios certificados excluidos dentro del ente territorial	text
3	tipo	Tipo de ente territorial, DPTO, ETC, MPIO	text

Fuente: Esta investigación.

Tabla 25. Diccionario de datos de la dimensión Municipio.

No.	ATRIBUTO	DESCRIPCIÓN	TIPO
PK	id_municipio	Llave primaria	serial
1	mun_cod_ente	Código DANE del municipio según divipo	text
2	mun_nombre	Nombre del municipio según divipo	text
3	mun_tipo	Tipo de ente territorial ETC, DPTO, MPIO	text

Fuente: Esta investigación.

Tabla 26. Diccionario de datos de la dimensión Tiempo.

No.	ATRIBUTO	DESCRIPCIÓN	TIPO
PK	id_tiempo	Llave primaria	serial
1	ano_prue	Año en que se realizó las pruebas.	text

Fuente: Esta investigación.

Tabla 27. Diccionario de datos de la tabla de hechos.

No.	ATRIBUTO	DESCRIPCIÓN	TIPO
PK	id_estudiante	Llave foránea	serial
PK	id_municipio	Llave foránea	integer
PK	id_ente	Llave foránea	integer
PK	id_establecimiento	Llave foránea	integer
PK	id_sede	Llave foránea	integer
PK	id_prueba	Llave foránea	integer
PK	id_tiempo	Llave foránea	integer
1	cod_consecutivo	Identificador del estudiante	text
2	numero_estu	Número de estudiantes	integer
3	weight1	Puntaje de la respectiva prueba1	numeric
4	weight2	Puntaje de la respectiva prueba2	numeric
5	promedio	Promedio entre las pruebas	numeric

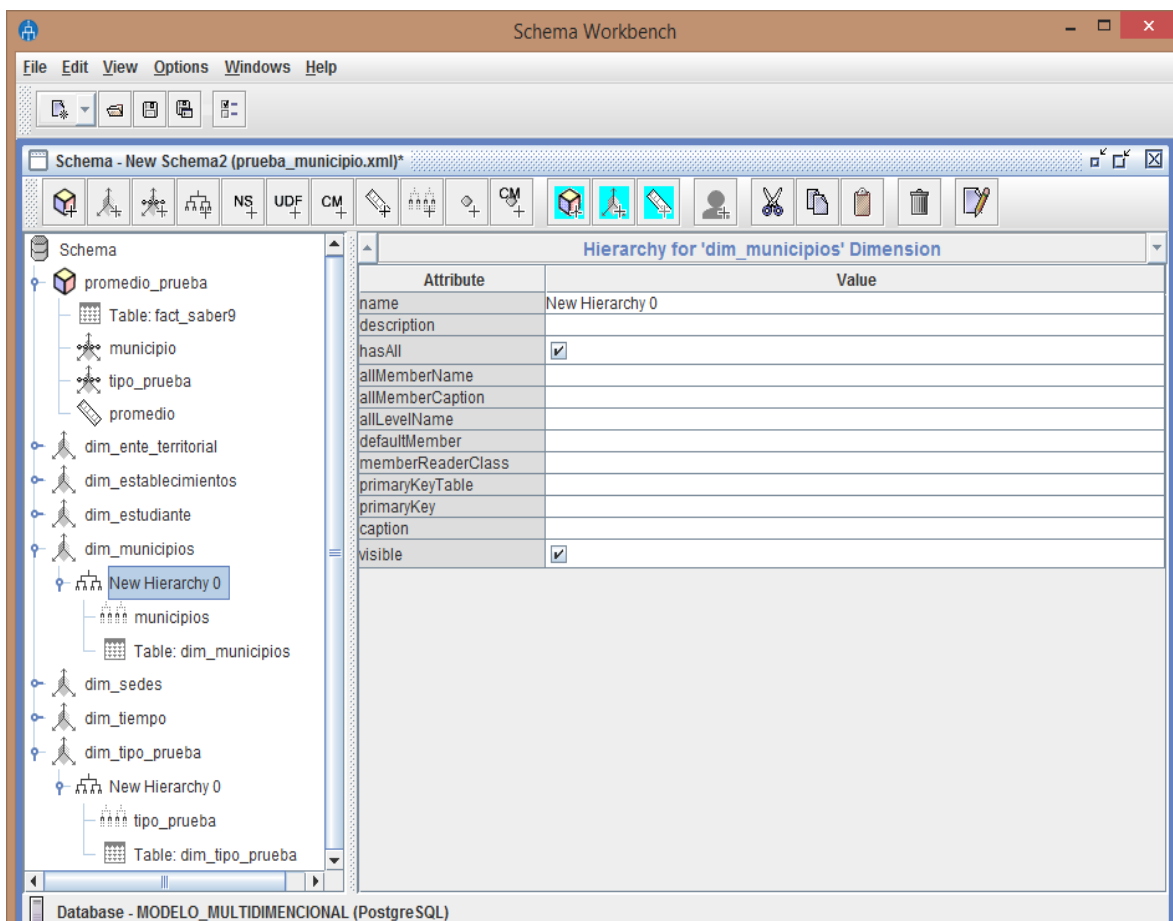
Fuente: Esta investigación.

1.6. Integración del mercado de datos con Pentaho

1.6.1. Cubo de datos. A continuación, en la figura 18, se presenta la configuración del cubo para las consultas OLAP, realizadas en la aplicación Schema Workbench de Mondrian, donde se tuvo en cuenta como primer paso configurar los parámetros de conexión con el Mercado de datos para poder crear el Schema el cual contendrá el cubo y las dimensiones.

Dentro del Schema creado se tuvo en cuenta el orden de los elementos para que su publicación sea de manera correcta, se creó una a una las dimensiones, configurando su nombre y su tabla de parámetros para acceder al Mercado de datos, por último, se creó el cubo y dentro del mismo se agregó la tabla de hechos y se configuro los parámetros correspondientes para que no existan fallos al momento de publicar el esquema.

Figura 18. Configuración del cubo para las consultas OLAP en Schema Workbench.



Fuente: Esta investigación.

En la figura 18 se observa la construcción del cubo de datos saber 9 se inició creando la conexión con el gestor postgres a la base de datos data_mart, como segundo paso se creó un schema con el nombre saber_9 a continuación se crearon las dimensiones dim_ente_territorial, dim_establecimientos, dim_estudiantes, dim_municipios, dim_sedes, dim_tiempo, dim_tipo_prueba, dim_copietas teniendo en cuenta que los siguientes pasos son igual para todas las anteriores dimensiones.

- Con click derecho sobre el schema se seleccionó Add dimension logrando crear las 8 dimensiones.
- Por cada dimensión se le asigna un nombre en este caso los mencionados anteriormente.
- El siguiente paso es darle un nombre a la jerarquía por cada dimensión.

- Como siguiente paso creamos las tablas haciendo click derecho sobre la jerarquía y seleccionamos Add Table para cada dimensión para lograr obtener los datos del gestor postgres.
- Ahora procedemos a crear los niveles de cada dimensión haciendo click derecho sobre cada jerarquía y posteriormente por cada nivel asignamos un nombre en este caso ente_territorial, establecimientos, estudiantes, municipios, sedes, tiempo, tipo_prueba, copietas, continuando seleccionamos la casilla namecolumn para traer nuestra base de datos de acuerdo a cada dimensión y en la casilla column seleccionamos nuestros niveles en este caso: id, id_establecimiento, id_estudiante, id_municipio, id_sede, id_tiempo,id_prueba.

Finalmente, creamos el cubo haciendo click derecho sobre el schema seleccionando la opción Add cube con el nombre cubo_saber9 a continuación haciendo click derecho sobre el cubo seleccionamos Add table para traer los datos de la tabla fact_saber9 para posteriormente usar las dimensiones ya creadas anteriormente, seleccionando la opción Add dimensión usage y por ultimo crear las medidas haciendo click derecho sobre el cubo seleccionando Add measure para obtener las medidas en este caso el promedio.

1.6.2. Operaciones OLAP

Las operaciones OLAP utilizadas para la visualización y análisis de los datos desde diferentes niveles de granularidad fueron las siguientes:

En la figura 19 se muestra el promedio por municipio y por competencias genéricas, siendo estos los datos originales.

Figura 19. Promedio por municipio y por competencias genéricas.

Prueba - Prueba	LENGUAJE_CIENCIAS	LENGUAJE_COMPETENCIAS	LENGUAJE_MATEMATICAS	MATEMATICAS_CIENCIAS	MATEMATICAS_COMPETENCIAS
Municipio	Promedio	Promedio	Promedio	Promedio	Promedio
ALDANA	1,524	1,661	1,568	1,5	1,8
CONTADERO	1,65	1,52	1,55	1,543	1,7
CORDOBA	1,576	1,662	1,673	1,574	1,654
CUASPUD	1,597	2,1	1,59	1,586	2,04
CUMBAL	1,651	1,631	1,627	1,628	1,595
FUNES	1,665	1,6	1,634	1,667	1,515
GUACHUCAL	1,567	1,9	1,657	1,574	1,691
GUALMATAN	1,616	1,57	1,615	1,637	1,6
ILES	1,545	1,612	1,692	1,611	1,943
IPIALES	1,646	1,665	1,709	1,658	1,699
POTOSI	1,616	1,87	1,688	1,561	1,848
PUERRES	1,611	2,6	1,651	1,614	2,8
PUPIALES	1,683	1,665	1,772	1,673	1,665

Fuente: Esta investigación.

En la figura 20 se muestra los datos cuando se aplica la operación OLAP, Drill-down para visualizar el promedio por establecimiento educativo, ente territorial y año de realización de la prueba.

Figura 20. Promedio por establecimiento educativo, ente territorial y año realización de la prueba.

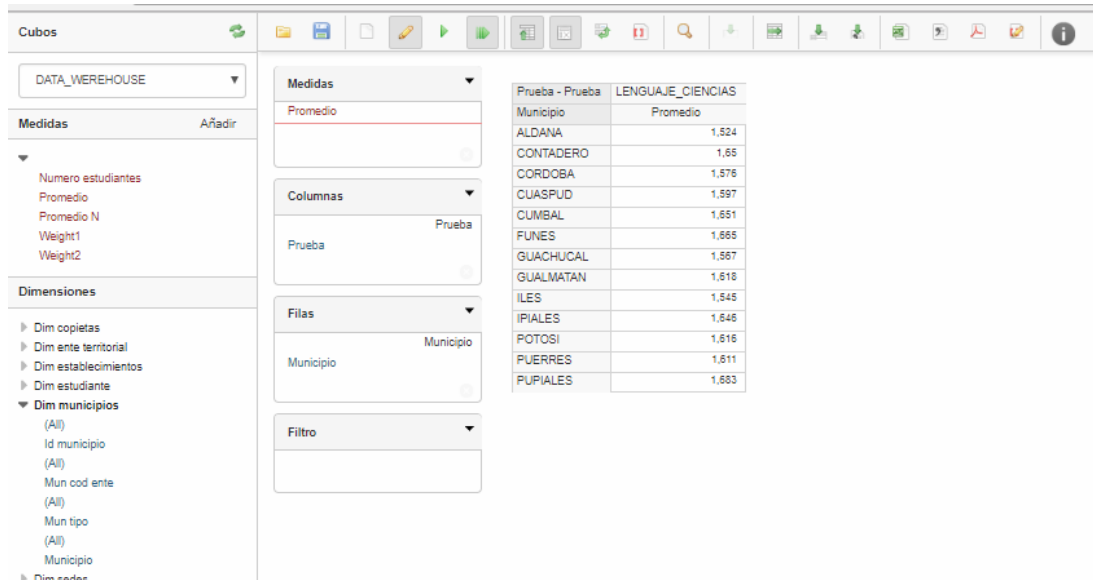
Información: 12.23 / 7 x 734 / 1.49s

Municipio	Establecimiento	Ente territorial	Prueba	Copia1	Año prueba	Promedio
ALDANA	INSTITUCION EDUCATIVA COMERCIAL SAN LUIS	NARIÑO	LENGUAJE_CIENCIAS	NO COPIA	2014	1,5
				NO COPIA	2016	1,5
			LENGUAJE_COMPETENCIAS	CON COPIA	2015	2,1
				NO COPIA	2015	2,1
			LENGUAJE_MATEMATICAS	NO COPIA	2014	1,5
				NO COPIA	2016	1,5
	MATEMATICAS_CIENCIAS	CON COPIA	2014	1,5		
		NO COPIA	2014	1,5		
	INSTITUCION EDUCATIVA NUESTRA SENORA DEL PILAR	NARIÑO	LENGUAJE_CIENCIAS	CON COPIA	2016	1,5
				NO COPIA	2016	1,5
			LENGUAJE_COMPETENCIAS	NO COPIA	2015	1,8
				CON COPIA	2015	1,7
LENGUAJE_MATEMATICAS			NO COPIA	2015	1,7	
			NO COPIA	2016	1,5	
MATEMATICAS_CIENCIAS	CON COPIA	2016	1,5			
	NO COPIA	2016	1,5			
MATEMATICAS_COMPETENCIAS	NARIÑO	CON COPIA	2015	1,8		
			NO COPIA	2015	1,8	
		CON COPIA	2015	1,8		
			NO COPIA	2015	1,8	
		CON COPIA	2015	1,8		
			NO COPIA	2015	1,8	
CONTADERO	INSTITUCION EDUCATIVA COLEGIO DEPTAL SAN CARLOS	NARIÑO	LENGUAJE_CIENCIAS	CON COPIA	2016	1,5
				NO COPIA	2016	1,5
			LENGUAJE_COMPETENCIAS	CON COPIA	2015	1,5
				NO COPIA	2015	1,5
			LENGUAJE_MATEMATICAS	NO COPIA	2016	1,5
				CON COPIA	2016	1,5
	MATEMATICAS_CIENCIAS	CON COPIA	2016	1,5		
		NO COPIA	2016	1,5		
	INSTITUCION EDUCATIVA LAS DELICIAS	NARIÑO	LENGUAJE_CIENCIAS	NO COPIA	2014	2,4
				NO COPIA	2016	1,8
			LENGUAJE_COMPETENCIAS	NO COPIA	2015	1,6
				NO COPIA	2015	1,7
LENGUAJE_MATEMATICAS			NO COPIA	2016	1,6	
			NO COPIA	2016	1,6	
MATEMATICAS_CIENCIAS	NO COPIA	2016	1,8			
	CON COPIA	2015	1,7			
MATEMATICAS_COMPETENCIAS	NARIÑO	NO COPIA	2015	1,7		
		NO COPIA	2015	1,7		
CORDOBA	CENTRO EDUCATIVO LLORENTE	NARIÑO	LENGUAJE_CIENCIAS	NO COPIA	2014	1,5
				NO COPIA	2016	2
			LENGUAJE_COMPETENCIAS	NO COPIA	2015	2
				NO COPIA	2015	2

Fuente: Esta investigación.

En la figura 21 se muestra los datos cuando se aplica la operación OLAP, Slice & Dice, para visualizar el promedio por municipio en la competencia genérica de lenguaje y ciencias.

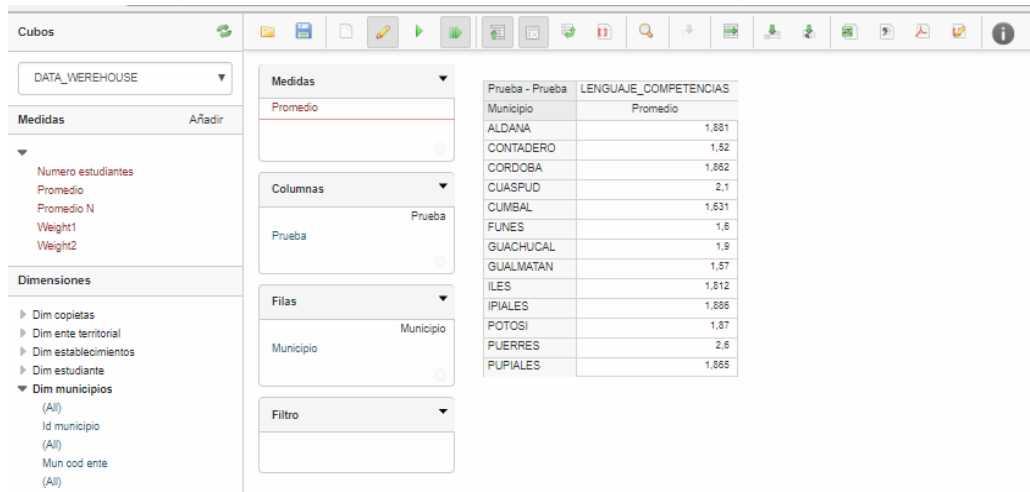
Figura 21. Promedio por municipio en lenguaje y ciencias.



Fuente: Esta investigación.

En la figura 22 se muestra los datos cuando se aplica la operación OLAP, Slice & Dice, para visualizar el promedio por municipio en la competencia genérica de lenguaje y competencias.

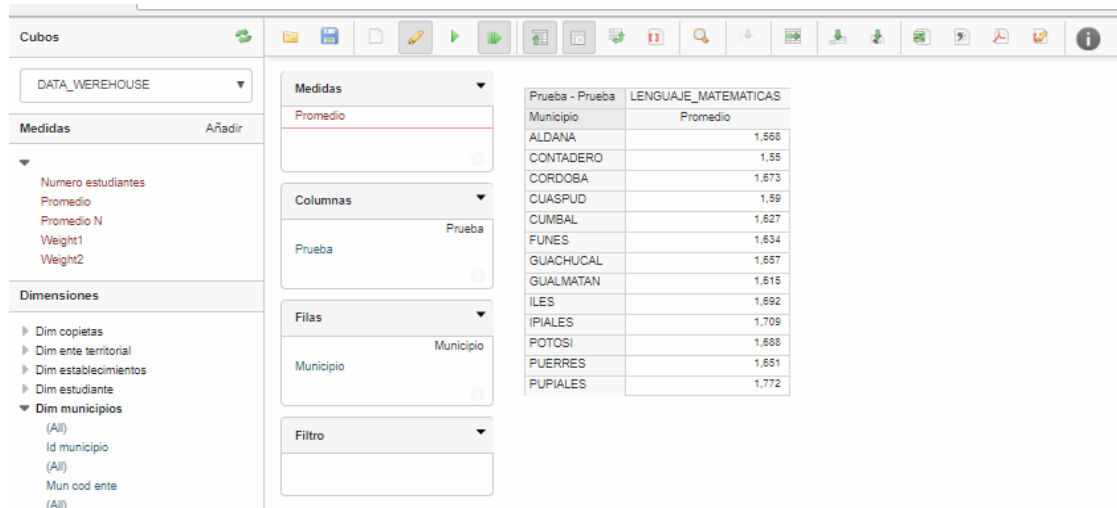
Figura 22. Promedio por municipio en lenguaje y competencias.



Fuente: Esta investigación.

En la figura 23 se muestra los datos cuando se aplica la operación OLAP, Slice & Dice, para visualizar el promedio por municipio en la competencia genérica de lenguaje y matemáticas.

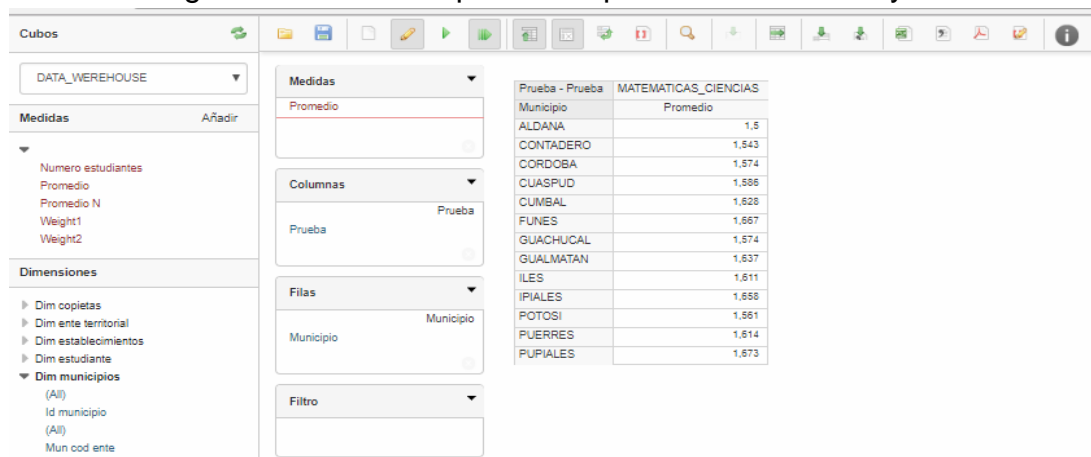
Figura 23. Promedio por municipio en lenguaje y matemáticas.



Fuente: Esta investigación.

En la figura 24 se muestra los datos cuando se aplica la operación OLAP, Slice & Dice, para visualizar el promedio por municipio en la competencia genérica de matemáticas y ciencias.

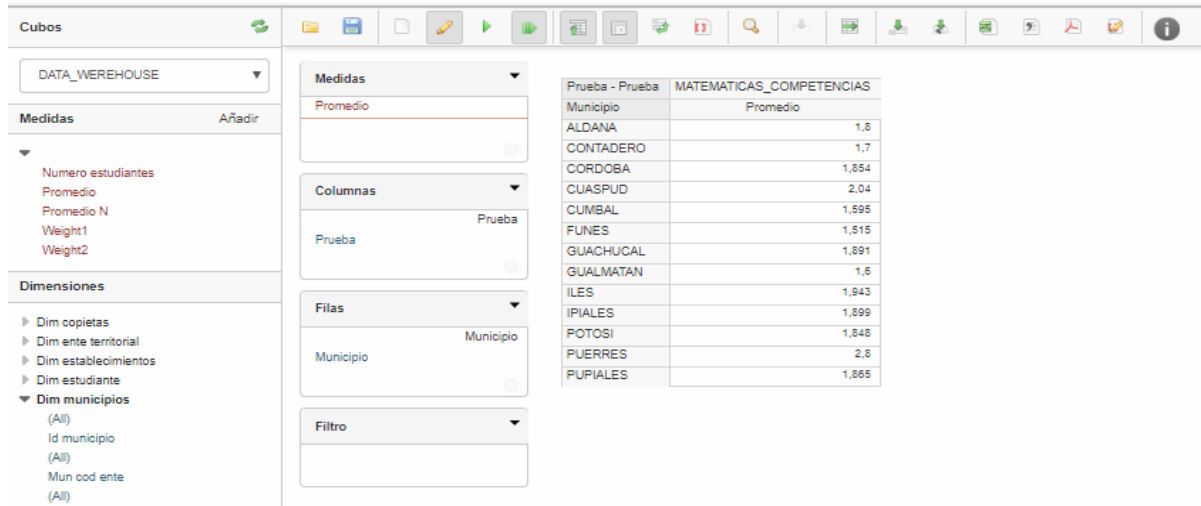
Figura 24. Promedio por municipio en matemáticas y ciencias.



Fuente: Esta investigación.

En la figura 25 se muestra los datos cuando se aplica la operación OLAP, Slice & Dice, para visualizar el promedio por municipio en la competencia genérica de matemáticas y competencias.

Figura 25. Promedio por municipio en matemáticas y competencias.



Fuente: Esta investigación.

1.7. Especificación y desarrollo de aplicaciones de BI

Para el diseño e implementación de los cubos que representan cada una de las medidas sobre los hechos se utilizó el programa libre bi-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 usando sentencias SQL y posteriormente se ajustaron las características de visualización. El procedimiento se muestra en las figuras, 19, 20,21

Figura 26. Pantalla de inicio Pentaho.



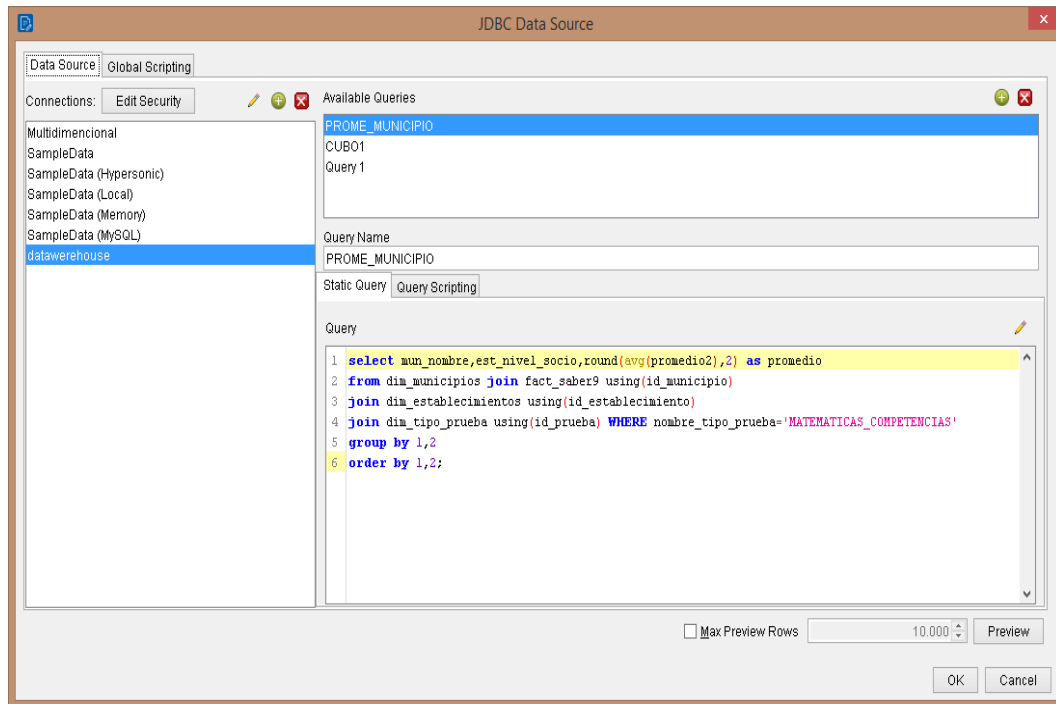
Fuente: Esta investigación.

Figura 27. Creación de reporte en Pentaho.



Fuente: Esta investigación.

Figura 28. Sentencia SQL para diseño de cubos.



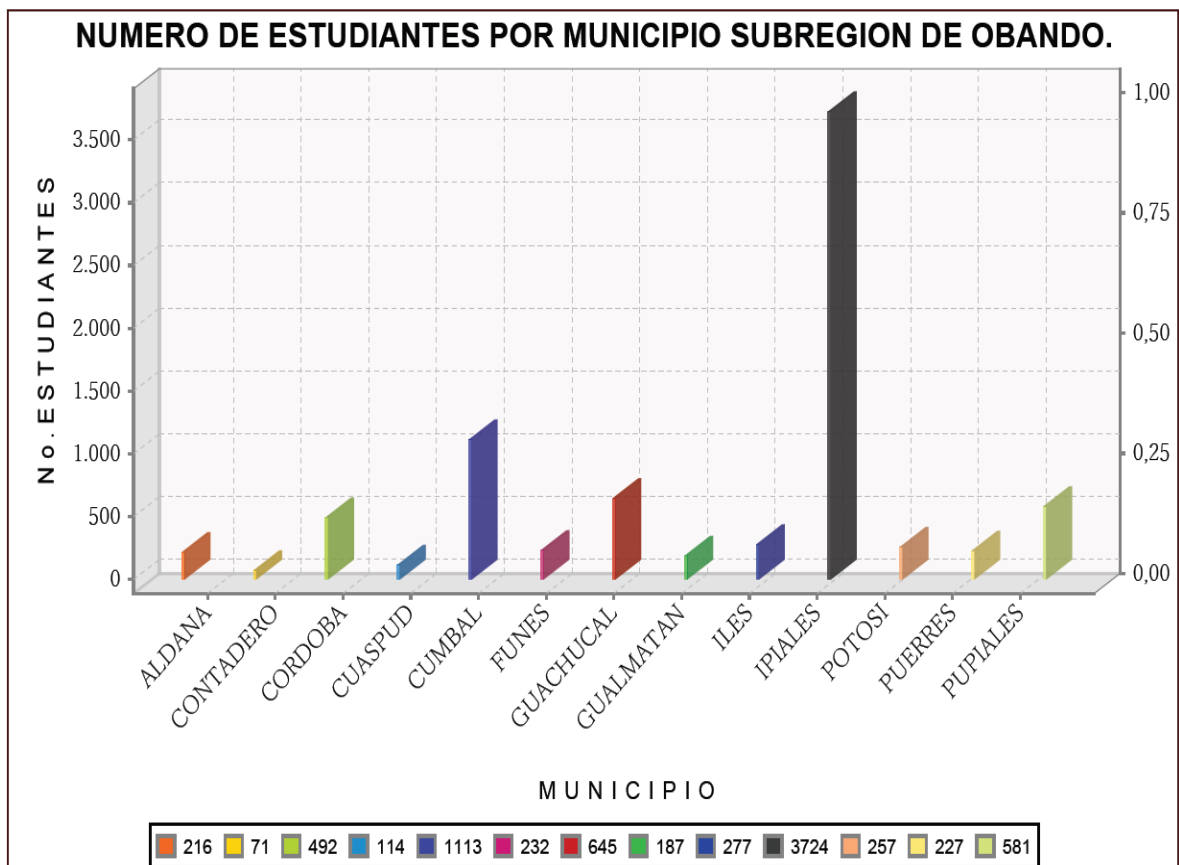
Fuente: Esta investigación.

2. RESULTADOS Y DISCUSIÓN

A partir del modelo multidimensional, se construyó un cubo para representar cada una de las medidas sobre los hechos.

1. ¿Cuántos estudiantes presentaron las Pruebas Saber 9° en el periodo comprendido entre 2014 y 2016? En la figura 22 se muestra los resultados utilizando Pentaho Report Designer, del número de estudiantes que presentaron la prueba saber 9° entre los años 2014 a 2016, clasificados por municipios, pertenecientes a la subregión de Obando.

Figura 29. Número de estudiantes que presentaron las Pruebas Saber 9° entre 2014 y 2016 clasificados por municipio.

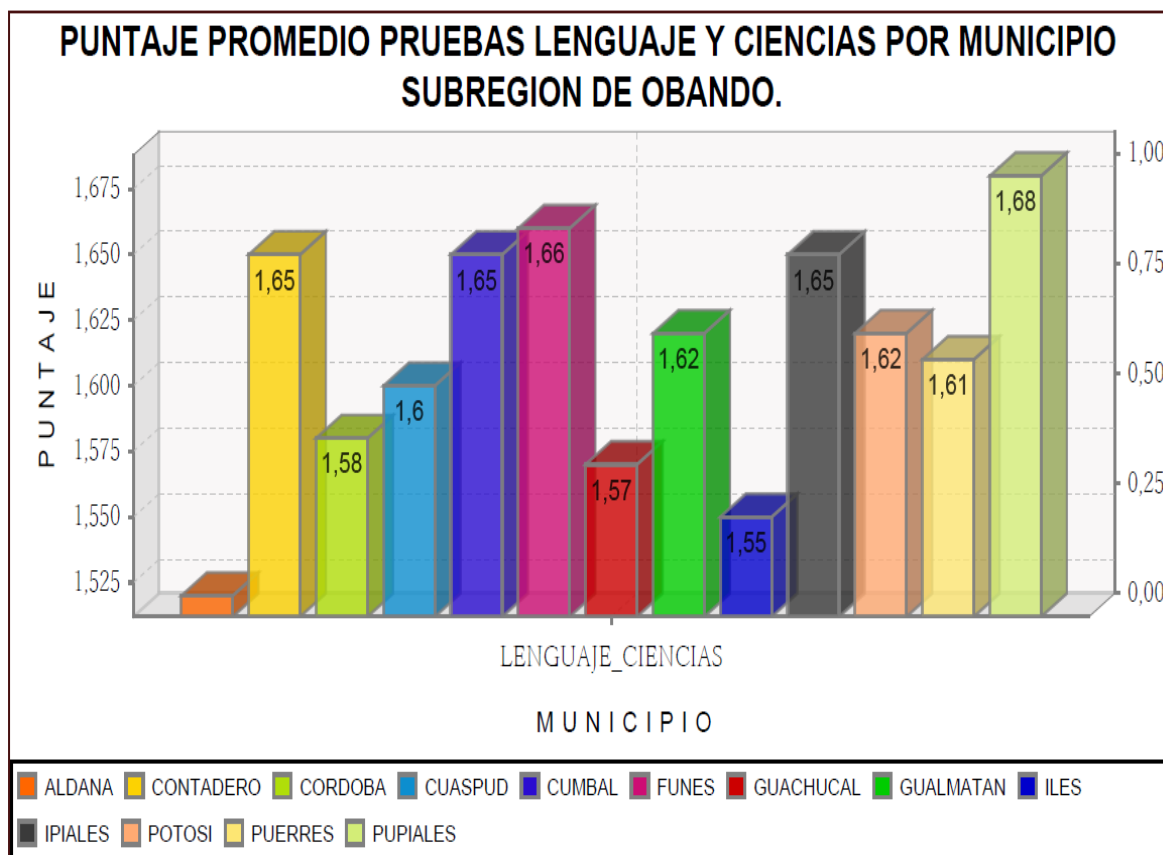


Fuente: Esta Investigación.

De la figura 29 se puede afirmar que, en la subregión de Obando, el municipio con mayor número de estudiantes que presentó la prueba saber 9, es Ipiales, esto debido a la cantidad de Instituciones Educativas existentes en dicho municipio, mientras que Contadero es el que menor cantidad de estudiantes se presentó a dicha prueba. Una razón para reafirmar dicha conclusión, es que el municipio de Contadero cuenta con una población de habitantes muy inferior frente a los demás municipios de la subregión de Obando.

¿Cuál fue el puntaje promedio de los resultados de lenguaje y ciencias por municipio en la subregión de Obando? En la figura 30 se muestra los resultados utilizando Pentaho Report Designer.

Figura 30. Puntaje promedio de los resultados de lenguaje y ciencias por municipio en la subregión de Obando.



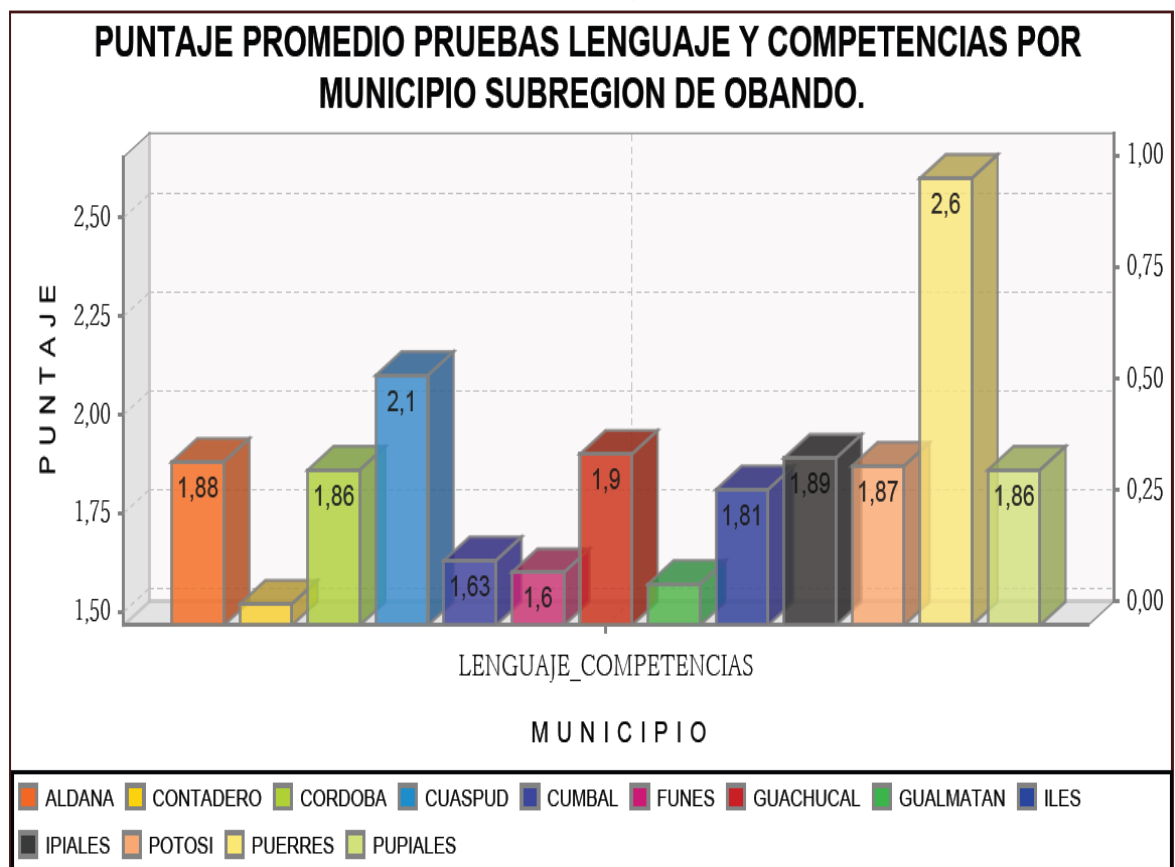
Fuente: Esta Investigación.

De la figura 30 se concluye, que el municipio de Pupiales se destaca con un puntaje promedio de 1.68 siendo este el mayor en lenguaje y ciencias, le siguen

Funes, Contadero, Cumbal e Ipiales con resultados favorables con respecto al primero. Mientras que el municipio de Aldana e Iles presentan un bajo rendimiento en relación al promedio que tienen los demás municipios, donde habría que tomar medidas de mejoramiento para estar a la par de los demás municipios.

¿Cuál fue el puntaje promedio de los resultados de lenguaje y competencias por municipio en la subregión de Obando? En la figura 31 se muestra los resultados utilizando Pentaho Report Designer

Figura 31. Puntaje promedio de los resultados de lenguaje y competencias por municipio en la subregión de Obando.



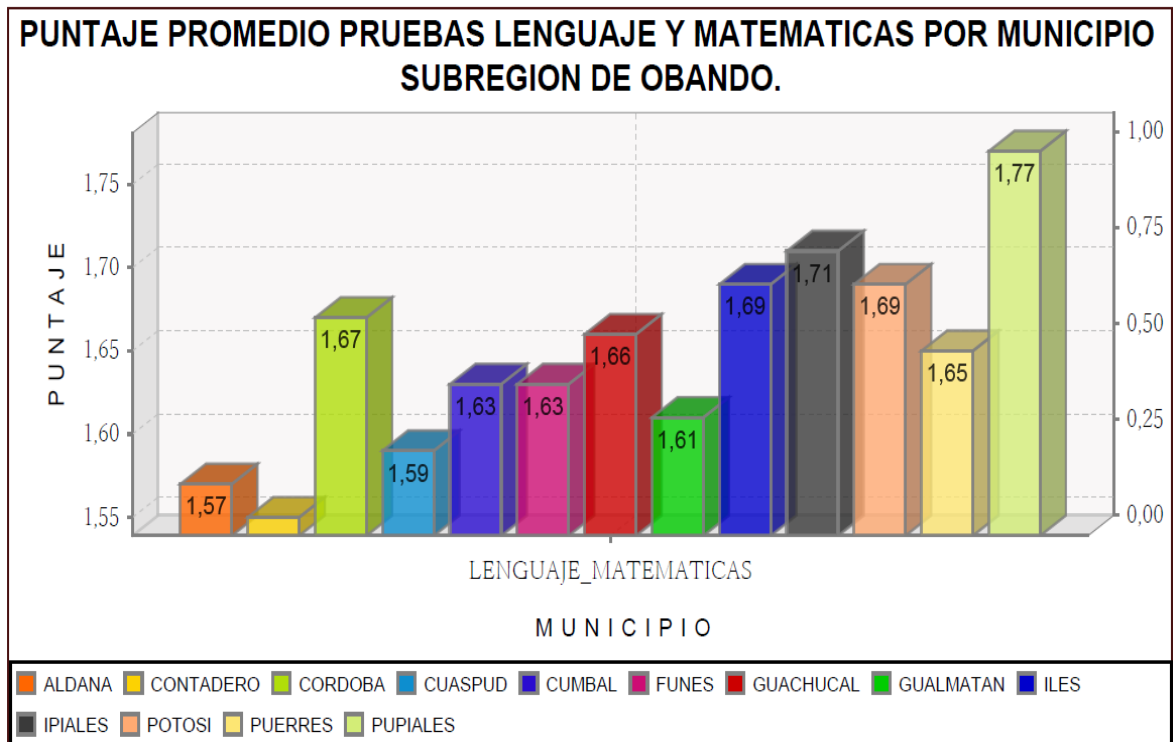
Fuente: Esta Investigación.

De la figura 31, observamos que la mayoría de municipios de la subregión de Obando, presenta dificultades en lenguaje y competencias, con puntajes que no superan el promedio de 2.0, a diferencia de los municipios de Puerres y Cuaspud con un promedio de 2,6 y 2,1 respectivamente. Lo cual resulta alarmante que tan

solo dos municipios de trece, su promedio sea mayor que 2.0, siendo este un motivo para realizar ajustes en la política educativa.

¿Cuál fue el puntaje promedio de los resultados de lenguaje y matemáticas por municipio en la subregión de Obando? En la figura 32 se muestra los resultados utilizando Pentaho Report Designer

Figura 32. Puntaje promedio de los resultados de lenguaje y matemáticas por municipio en la subregión de Obando.

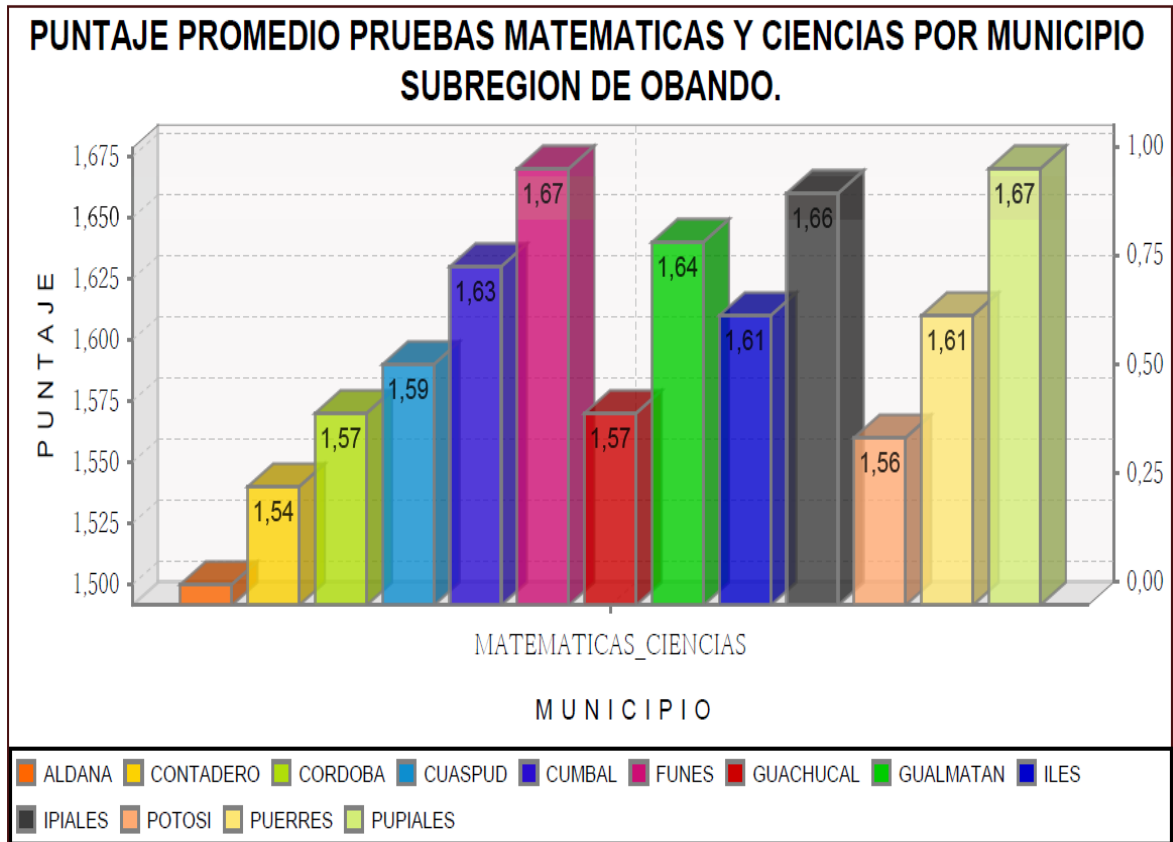


Fuente: Esta Investigación.

De la figura 32 se puede afirmar que, en lenguaje y matemáticas, los promedios obtenidos por los municipios de la subregión de Obando presentan un comportamiento casi similar. En el caso extremo se encuentran los municipios de Contadero, Aldana y Cuaspud donde se debería buscar alternativas de mejoramiento lo más pronto posible para no alejarse de los demás municipios, aunque cabe aclarar que el promedio en general no es muy bueno ya que ningún municipio supera el promedio de 2.0

¿Cuál fue el puntaje promedio de los resultados de matemáticas y ciencias por municipio en la subregión de Obando? En la figura 33 se muestra los resultados utilizando Pentaho Report Designer

Figura 33. Puntaje promedio de los resultados de matemáticas y ciencias por municipio en la subregión de Obando.

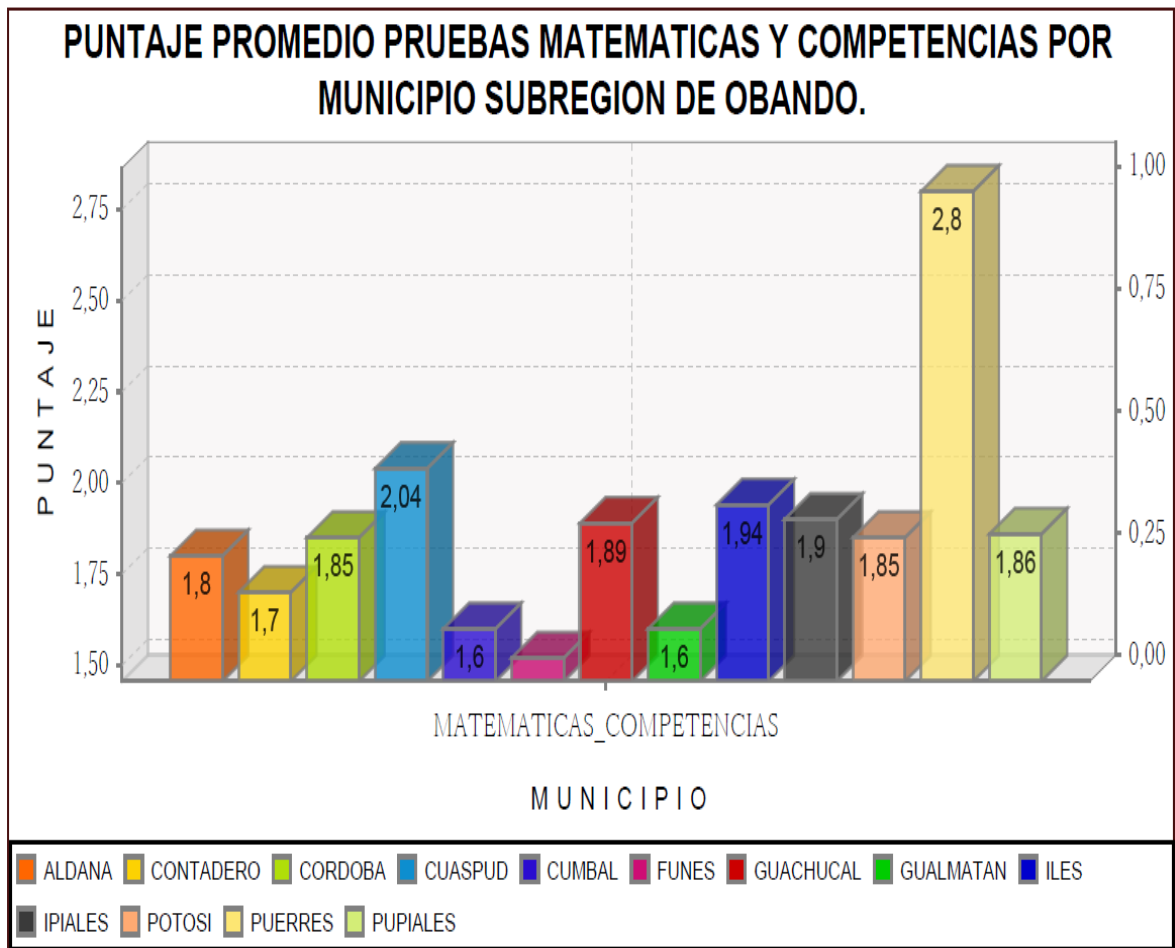


Fuente: Esta Investigación.

De acuerdo a la figura 33 se concluye que, en matemáticas y ciencias los promedios de los municipios de la subregión de Obando son relativamente bajos y no logran superar un promedio de 2.0. Siendo esta una medida para ajustar las políticas educativas en todas las Instituciones Educativas de la subregión, con el fin de fortalecer y contrarrestar las debilidades que se presentan en matemáticas y ciencias. Caso particular es el que presenta el municipio de Aldana el cual se lo podría considerar como alerta máxima de mejoramiento.

¿Cuál fue el puntaje promedio de los resultados de matemáticas y competencias por municipio en la subregión de Obando? En la figura 34 se muestra los resultados utilizando Pentaho Report Designer.

Figura 34. Puntaje promedio de los resultados de matemáticas y competencias por municipio en la subregión de Obando.

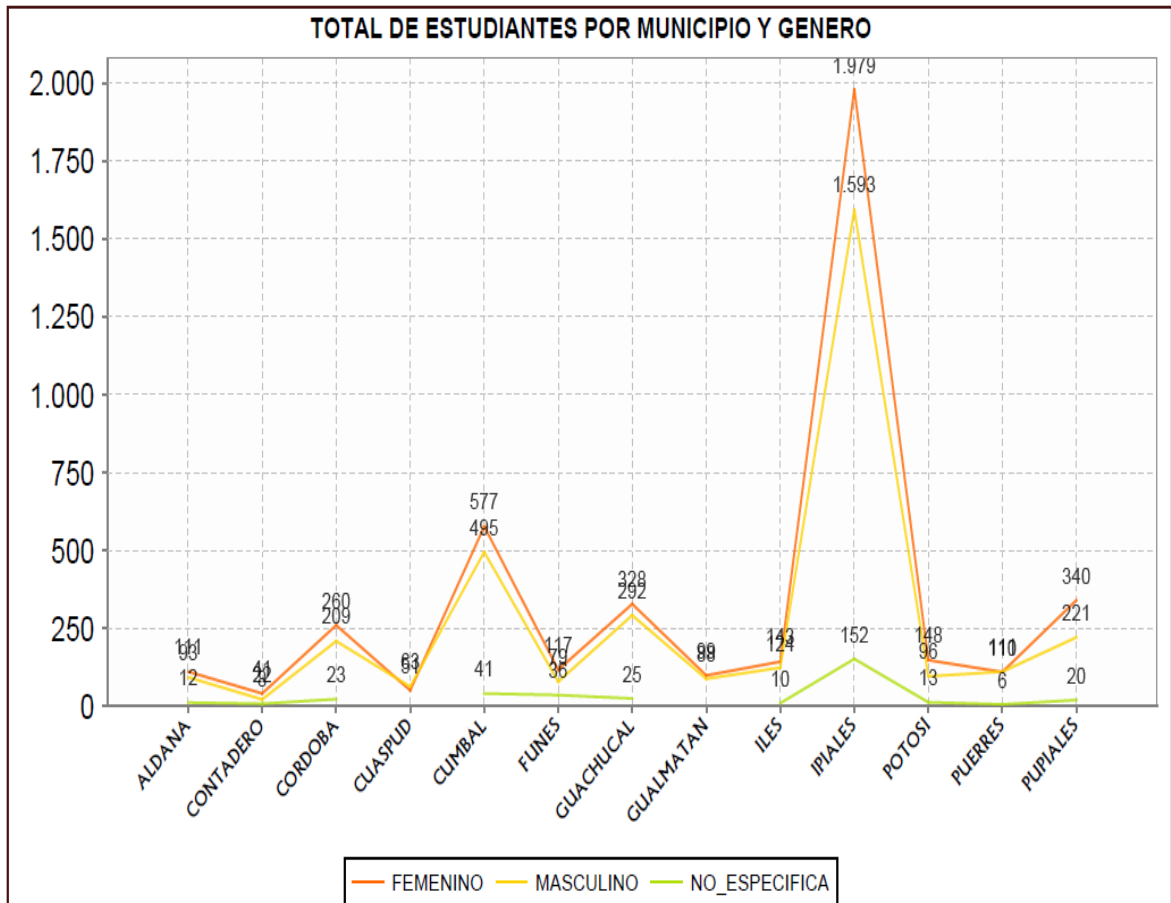


Fuente: Esta Investigación.

En la figura 34 notoriamente se observa que el municipio de Puerres presenta un promedio significativamente alto en matemáticas y competencias con respecto a los demás municipios de la subregión de Obando, de igual forma Cuaspud logra sobrepasar el promedio de 2.0 mientras que el resto de municipios se mantienen con un promedio que casi logra alcanzar el 2.0 como el caso de Iles e Ipiales, que con políticas de mejoramiento pueden lograr alcanzar el objetivo.

¿Cuál es el número de estudiantes por municipio y género que presentaron las pruebas saber 9° comprendido entre los años 2014 – 2016? En la figura 35 se muestra los resultados utilizando Pentaho Report Designer.

Figura 35. Número de estudiantes por municipio y género que presentaron las pruebas saber 9° comprendido entre los años 2014 a 2016.

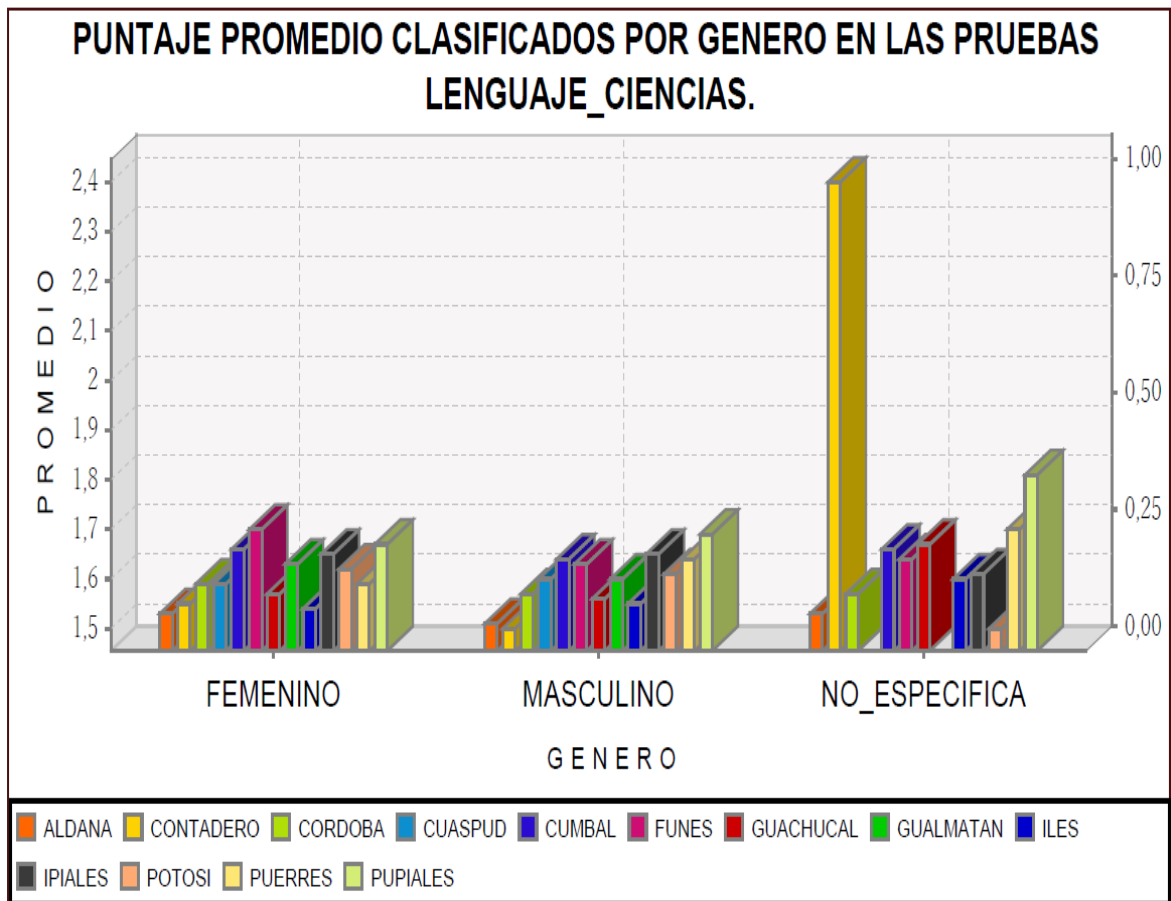


Fuente: Esta Investigación.

En la figura 35 podemos observar que existe una mínima diferencia en cuanto al número de estudiantes que presentaron la prueba saber 9° clasificados por género. Si bien el género 'Femenino' aventaja al género 'Masculino' en todos los municipios de la subregión de Obando la diferencia no sobrepasa de los cien estudiantes, salvo en el caso de los municipios de Ipiales y Pupiales donde la diferencia es más notoria siendo dominado por el género 'Femenino'. Cabe resaltar que la variable 'No_Especifica' tiene un comportamiento mínimo frente a las variables mencionadas anteriormente.

¿Cuál es el puntaje promedio de los resultados por municipio, clasificados por género en las áreas de lenguaje y ciencias? En la figura 36 se muestra los resultados utilizando Pentaho Report Designer

Figura 36. Puntaje promedio de los resultados por municipio, clasificado por género en las áreas de lenguaje y ciencias.



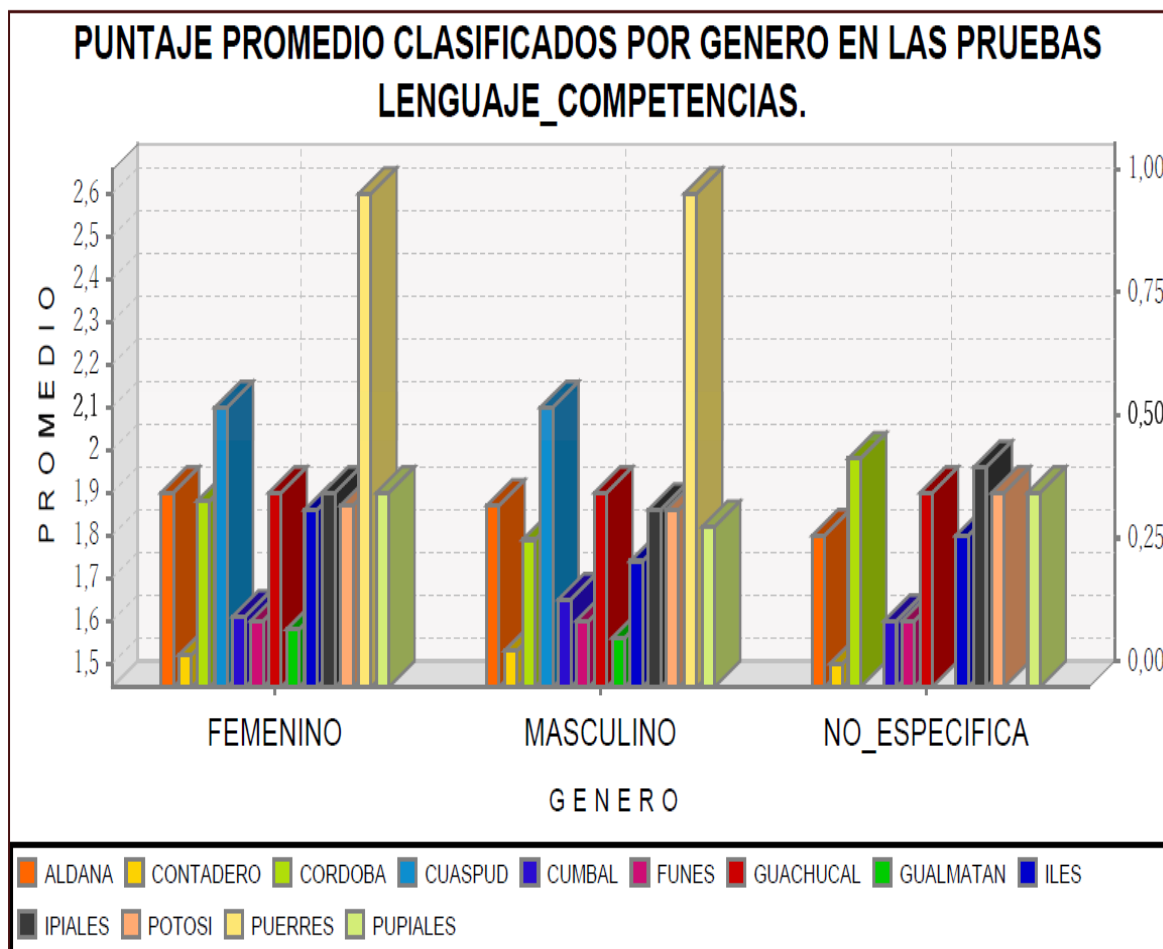
Fuente: Esta Investigación

De la figura 36 se aprecia que, el puntaje promedio de los municipios tanto en el género 'Femenino' como 'Masculino' en lenguaje y ciencias, presentan un comportamiento casi similar, siendo en el género Femenino, el municipio de Funes con el promedio mayor, mientras que Aldana, Contadero e Iles presentan un promedio bajo. Por otra parte, en el género masculino, es el municipio de PuiALES quien presenta un promedio mayor frente a los demás municipios, sin embargo, los promedios no sobrepasan el promedio de 2.0.

Caso contrario ocurre en la variable 'No_Especifica' donde se destaca el municipio de Contadero con un promedio significativamente alto con respecto a los demás municipios de esta misma variable y también con los de género femenino y masculino. A demás en esta variable los municipios de Cuaspud y Gualmatan la cantidad de estudiantes es cero.

¿Cuál es el puntaje promedio de los resultados por municipio, clasificados por género en las áreas de lenguaje y competencias? En la figura 37 se muestra los resultados utilizando Pentaho Report Designer.

Figura 37. Puntaje promedio de los resultados por municipio, clasificado por género en las áreas de lenguaje y competencias.

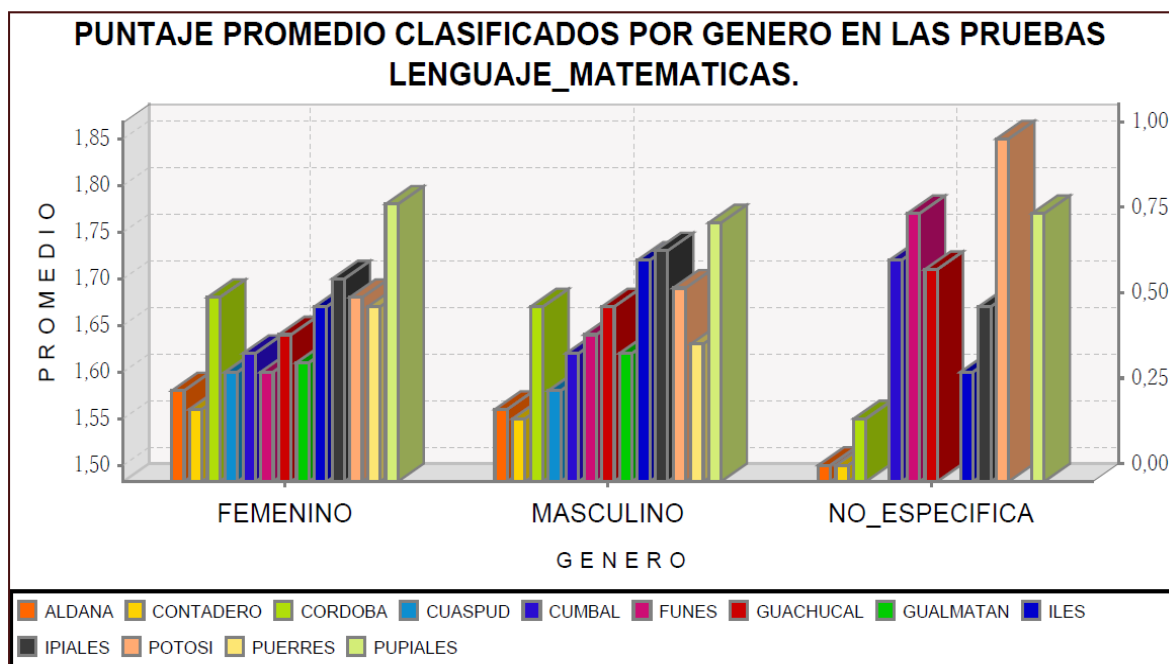


Fuente: Esta Investigación.

De la figura 37, se observa el promedio que obtuvieron los estudiantes de los municipios de la subregión de Obando en lenguaje y competencias, siendo estos promedios satisfactorios tanto en el género Femenino como Masculino, ya que algunos municipios alcanzan un promedio mayor a 2.0. En el género Femenino se destaca el municipio de Puerres con un buen promedio de 2.6, le sigue Cuaspud con 2.1 y más atrás están Aldana, Guachucal, Ipiates y Pupiales y el promedio menor lo tiene Contadero. De igual forma en el género masculino se mantienen los mismos municipios con el mayor promedio Puerres y Cuaspud y a Contadero con el menor promedio. En la variable 'No_Especifica' los municipios de Cuaspud, Gualmatan y Puerres la cantidad de estudiantes es cero.

¿Cuál es el puntaje promedio de los resultados por municipio, clasificados por género en las áreas de lenguaje y matemáticas? En la figura 38 se muestra los resultados utilizando Pentaho Report Designer

Figura 38. Puntaje promedio de los resultados por municipio, clasificado por género en las áreas de lenguaje y matemáticas.



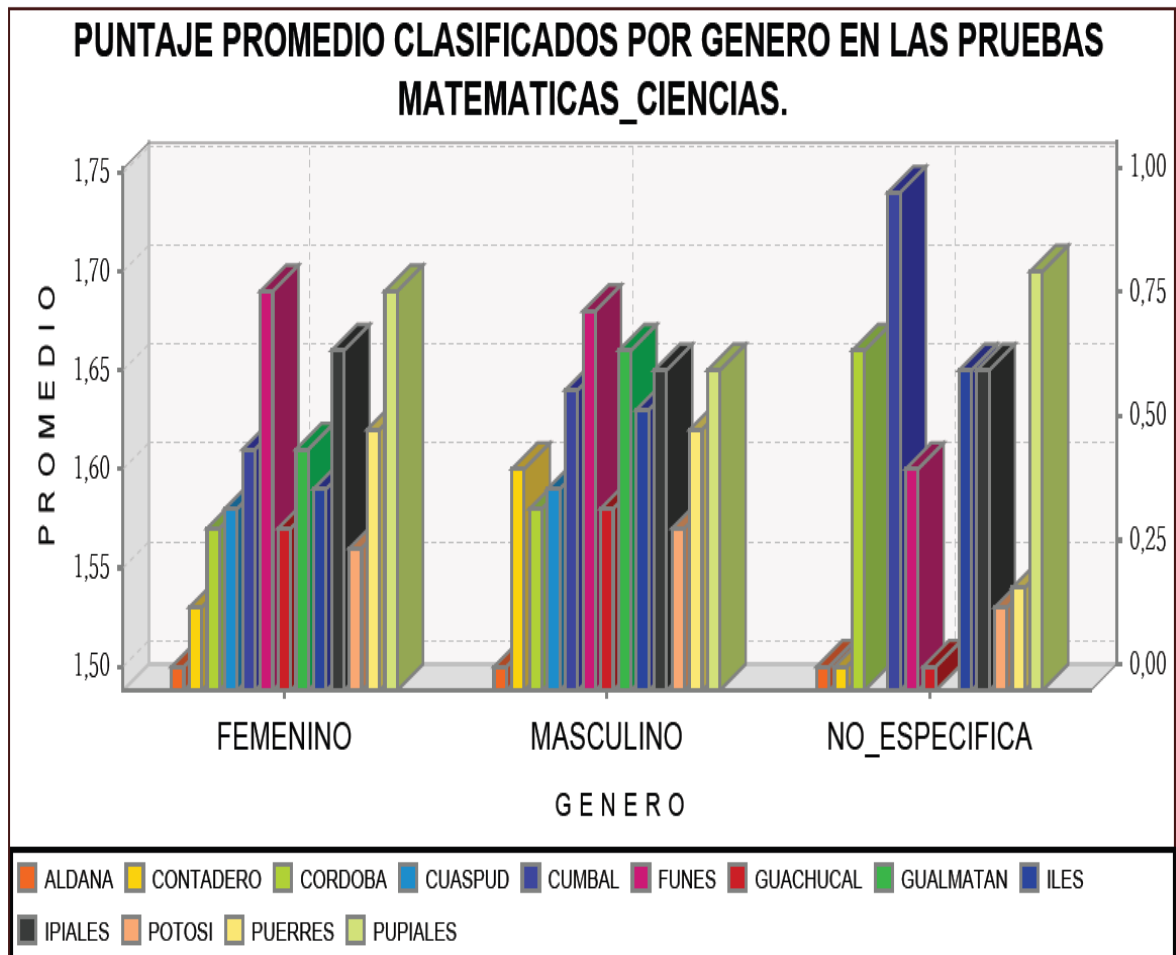
Fuente: Esta Investigación

De la figura 38, los promedios nuevamente bajan en lenguaje y matemáticas, ya que ningún municipio de la subregión de Obando logra obtener un promedio igual o mayor a 2.0 tanto en el género femenino, masculino como en la variable

'No_Especifica'. En el género Femenino, se destacan los municipios de Pupiales, Ipiiales y Córdoba, aunque los promedios no son para nada favorables, mientras que en el género Masculino se presenta el mismo comportamiento, determinando así que es necesario tomar medidas de mejoramiento para mejorar la calidad educativa. Por último, en la variable 'No_Especifica' los municipios de Cuaspud, Gualmatan y Puerres el número de estudiantes es cero.

¿Cuál es el puntaje promedio de los resultados por municipio, clasificados por género en las áreas de matemáticas y ciencias? En la figura 39 se muestra los resultados utilizando Pentaho Report Designer

Figura 39. Puntaje promedio de los resultados por municipio, clasificado por género en las áreas de matemáticas y ciencias.



Fuente: Esta Investigación

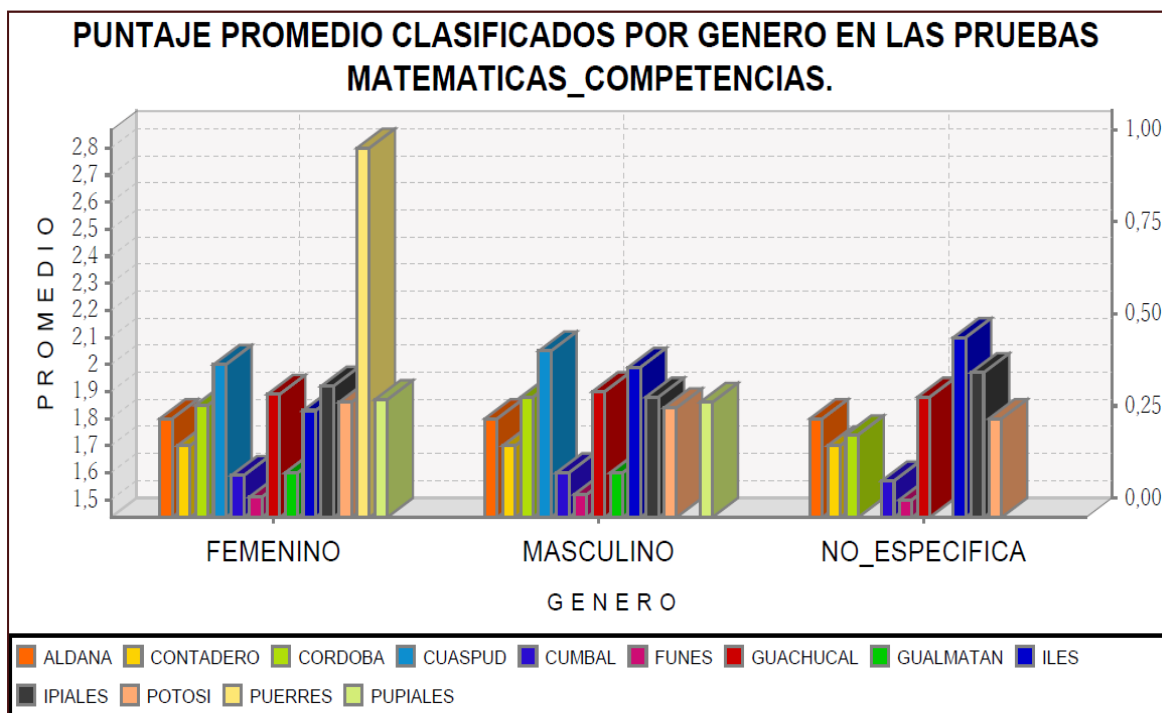
De acuerdo a la figura 39 se puede observar que, el municipio de Aldana es el que presenta el menor promedio en matemáticas y ciencias, tanto en el género femenino como en el masculino, de igual forma en la variable 'No_Especifica'. En esta última también se encuentran los municipios de Contadero y Guachucal con promedios relativamente bajos.

En el género femenino se destacan los municipios de Funes y Pupiales los cuales presentan promedios de 1.7 pero esto no significa que no se deba replantear las políticas educativas, al contrario, es en todos los municipios que se necesita tomar medidas de mejoramiento, ya que los promedios no alcanzan el 2.0.

En la variable 'No_Especifica' los municipios de Cuaspud y Gualmatan el número de estudiantes es cero.

¿Cuál es el puntaje promedio de los resultados por municipio, clasificados por género en las áreas de matemáticas y competencias? En la figura 40 se muestra los resultados utilizando Pentaho Report Designer

Figura 40. Puntaje promedio de los resultados por municipio, clasificado por género en las áreas de matemáticas y competencias.



Fuente: Esta Investigación

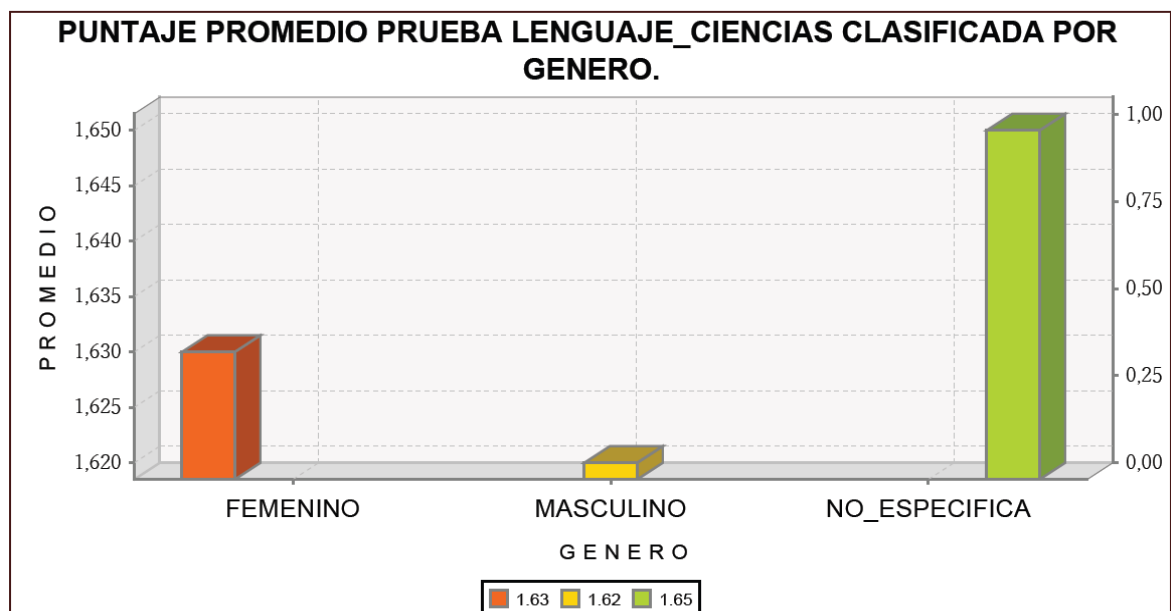
De la figura 40 se puede concluir que, los municipios de la subregión de Obando presentan promedios significativamente altos en matemáticas y competencias tanto en el género femenino como en el masculino.

En el género femenino se aprecia que el municipio de Puerres obtiene el mayor promedio y le siguen los municipios de Cuaspud, Guachucal, Ipiales y Pupiales, también se observa los buenos resultados que presenta el municipio de Potosí, caso contrario sucede con Cumbal, Funes y Gualmatan donde se deberían tomar medidas de mejoramiento para estar a la par del resto de los municipios que presentan promedios favorables.

En la variable 'No_Especifica' los municipios de Cuaspud, Gualmatan, Puerres y Pupiales el número de estudiantes es cero.

¿Cuál es el puntaje promedio de la prueba de lenguaje y ciencias clasificada por género? En la figura 41 se muestra los resultados utilizando Pentaho Report Designer

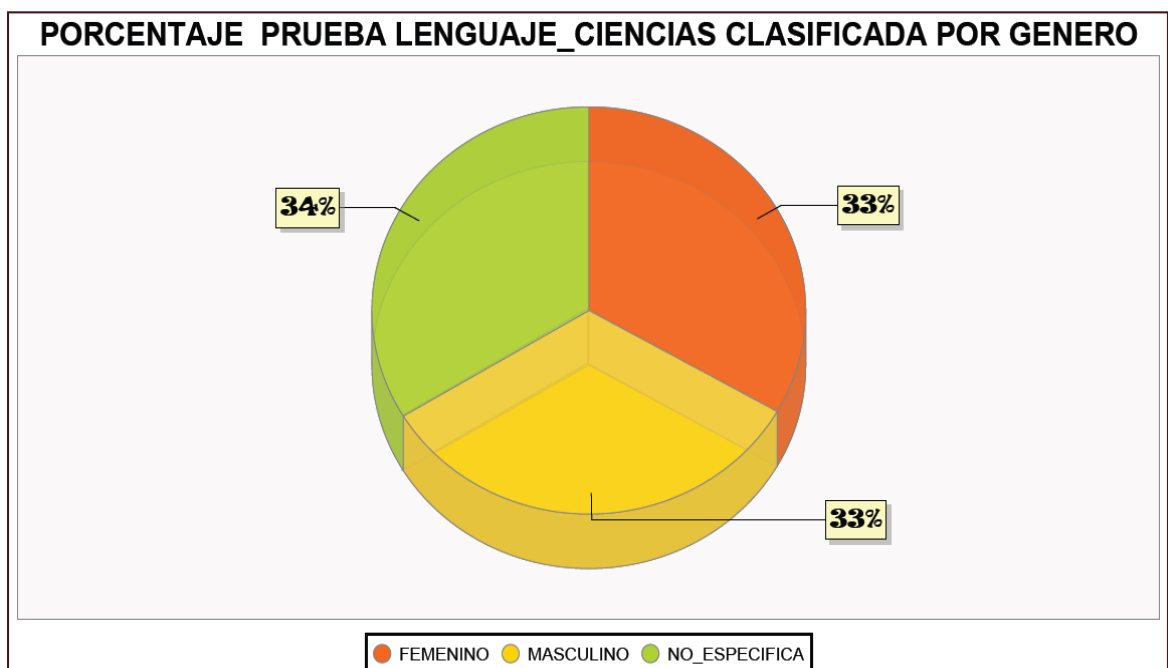
Figura 41. Puntaje promedio de la prueba de lenguaje y ciencias clasificada por género.



Fuente: Esta investigación.

De la figura 41, se concluye que la variable 'No_Especifica', obtuvo el promedio mayor en lenguaje y ciencias, 1.65. De igual forma cabe resaltar que la diferencia con respecto al género 'Femenino' es mínima, 1.63 y finalmente encontramos el género 'Masculino' con un promedio de 1.62. Por otra parte, en la figura 42 se muestra los mismos resultados utilizando Pentaho Report Designer para detallar el porcentaje equivalente a los promedios obtenidos por el género 'Femenino', 'Masculino' y la variable 'No_Especifica'

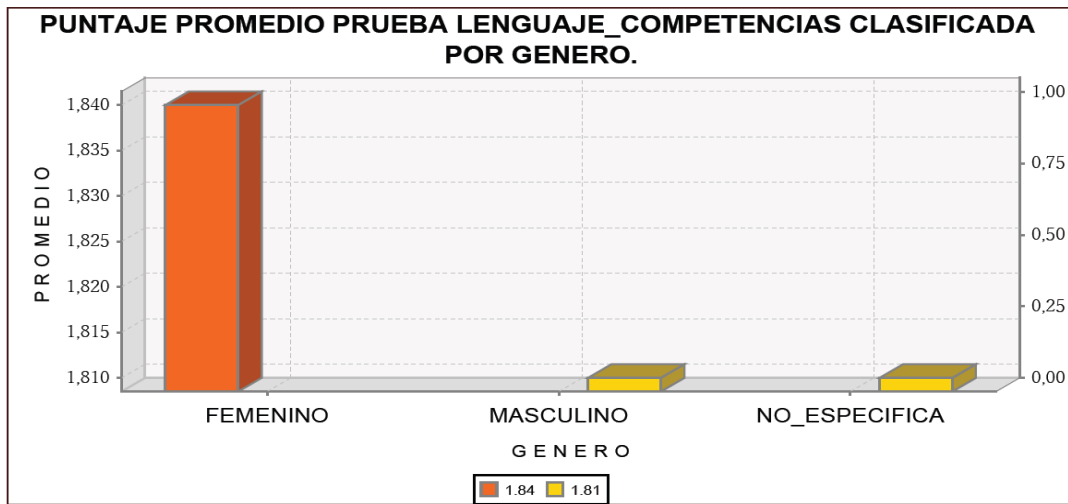
Figura 42. Porcentaje de estudiantes que presentaron la prueba de lenguaje y ciencias, clasificada por género.



Fuente: Esta investigación.

¿Cuál es el puntaje promedio de la prueba de lenguaje y competencias clasificada por género? En la figura 43 se muestra los resultados utilizando Pentaho Report Designer

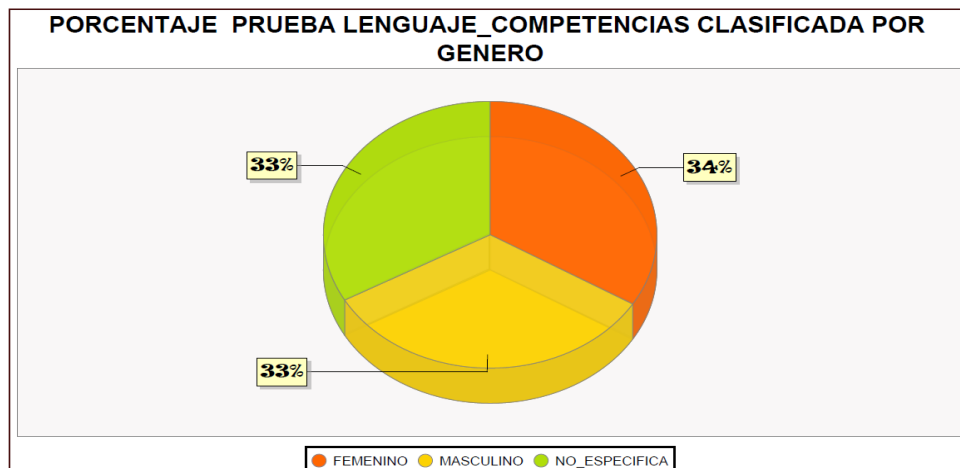
Figura 43. Puntaje promedio de la prueba de lenguaje y competencias clasificada por género.



Fuente: Esta investigación.

De acuerdo a la figura 43, se concluye que el género femenino sobresale con un promedio de 1.84, en lenguaje y competencias, siendo este el promedio mayor sobre el género masculino y la variable No_Especifica los cuales presentan el mismo promedio de 1.81. Por otra parte, en la figura 44 se muestra los mismos resultados utilizando Pentaho Report Designer para detallar el porcentaje equivalente a los promedios obtenidos por el género 'Femenino', 'Masculino' y la variable 'No_Especifica'

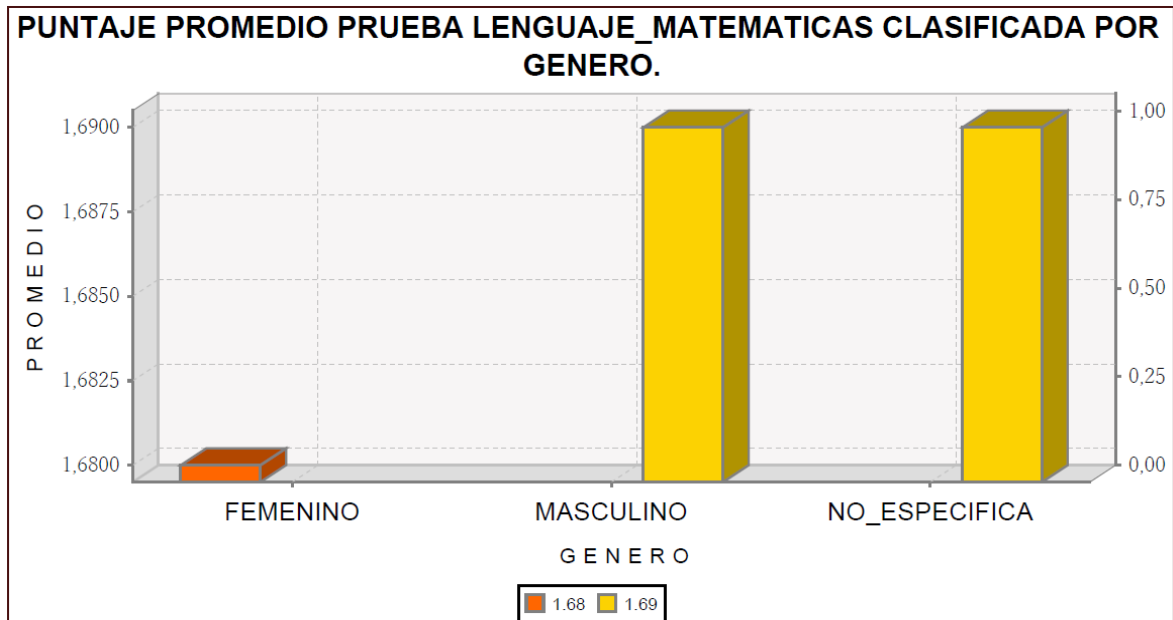
Figura 44. Porcentaje de estudiantes que presentaron la prueba de lenguaje y competencias, clasificada por género.



Fuente: Esta investigación.

¿Cuál es el puntaje promedio de la prueba de lenguaje y matemáticas clasificada por género? En la figura 45 se muestra los resultados utilizando Pentaho Report Designer

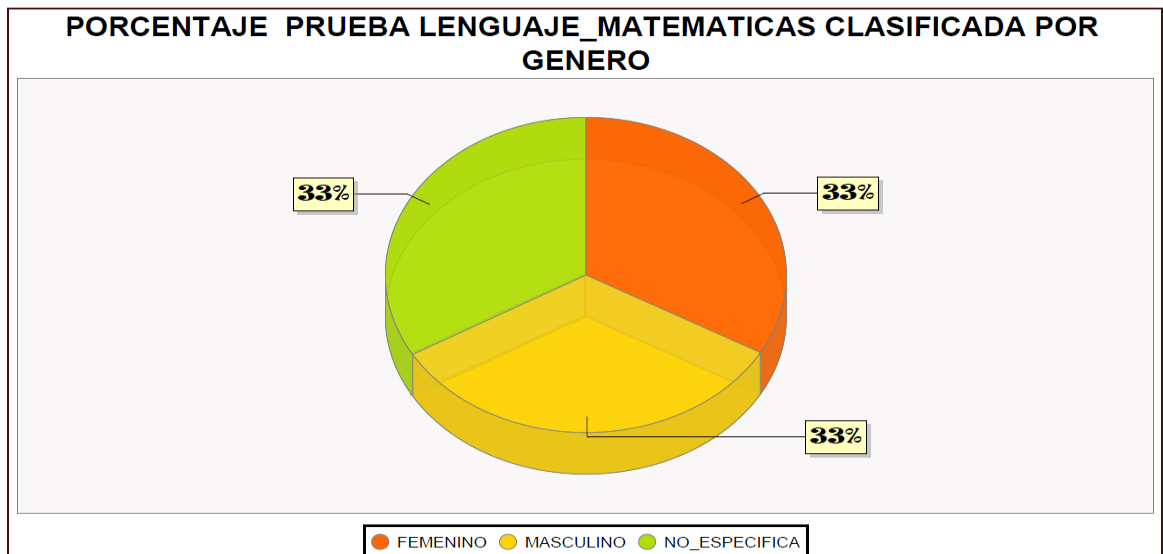
Figura 45. Puntaje promedio de la prueba de lenguaje y matemáticas clasificada por género.



Fuente: Esta investigación.

De acuerdo a la figura 45, el género masculino y la variable No_Especificia se destacan con el mismo promedio, 1.69 sobre el género femenino que presenta un promedio de 1.68 donde la diferencia es mínima en lenguaje y matemáticas. Por otra parte, en la figura 46 se muestra los mismos resultados utilizando Pentaho Report Designer para detallar el porcentaje equivalente a los promedios obtenidos por el género 'Femenino', 'Masculino' y la variable 'No_Especificia'

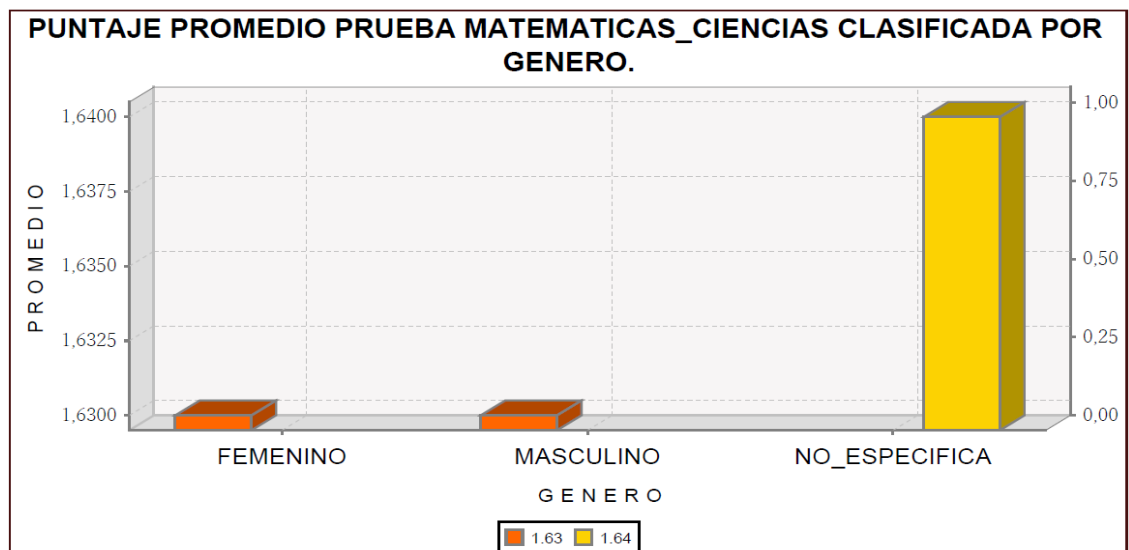
Figura 46. Porcentaje de estudiantes que presentaron la prueba de lenguaje y matemáticas, clasificada por género.



Fuente: Esta investigación.

¿Cuál es el puntaje promedio de la prueba de matemáticas y ciencias clasificada por género? En la figura 47 se muestra los resultados utilizando Pentaho Report Designer

Figura 47. Porcentaje de estudiantes que presentaron la prueba de lenguaje y matemáticas, clasificada por género.

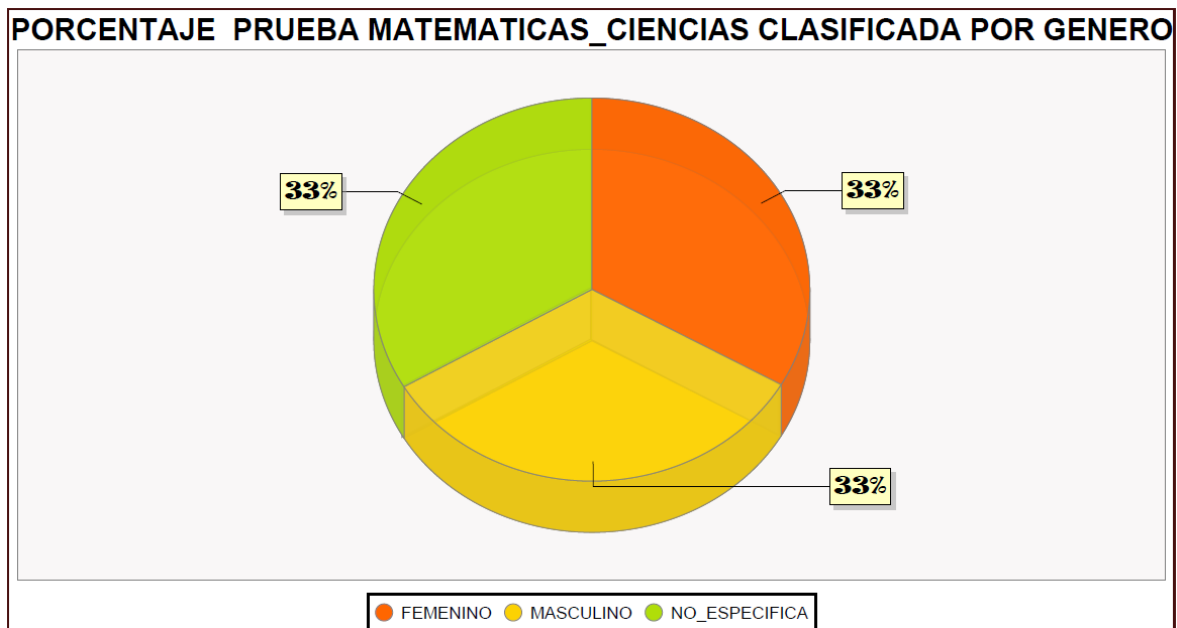


Fuente: Esta investigación.

De la figura 47, nuevamente la variable No_Especifica se destaca con el mayor promedio de 1.64 con una mínima diferencia sobre el género femenino y masculino que presentan el mismo promedio, 163, en matemáticas y ciencias.

Por otra parte, en la figura 48 se muestra los mismos resultados utilizando Pentaho Report Designer para detallar el porcentaje equivalente a los promedios obtenidos por el género 'Femenino', 'Masculino' y la variable 'No_Especifica'

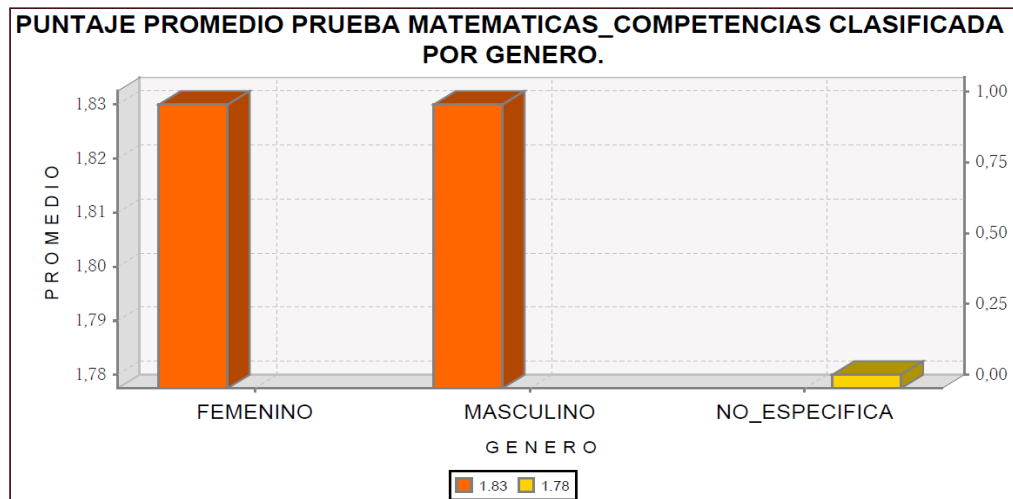
Figura 48. Porcentaje de estudiantes que presentaron la prueba de matemáticas y ciencias, clasificada por género.



Fuente: Esta investigación.

¿Cuál es el puntaje promedio de la prueba de matemáticas y competencias por género? En la figura 49 se muestra los resultados utilizando Pentaho Report Designer

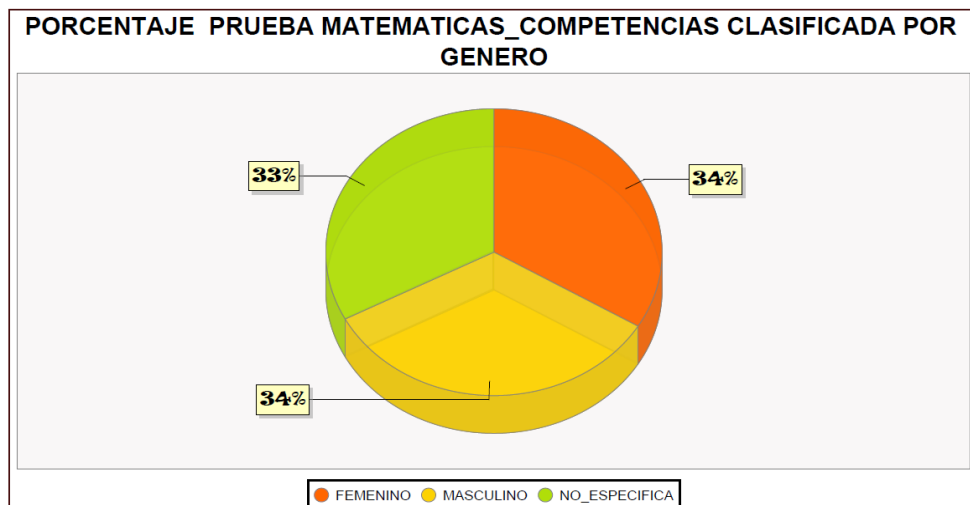
Figura 49. Puntaje promedio de la prueba de matemáticas y competencias clasificada por género.



Fuente: Esta investigación.

De la figura 49, tanto el género femenino como masculino presentan promedios similares de 1.83 siendo estos los mayores con respecto a No_Especificia con un promedio de 1.78 en matemáticas y competencias. Por otra parte, en la figura 50 se muestra los mismos resultados utilizando Pentaho Report Designer para detallar el porcentaje equivalente a los promedios obtenidos por el género 'Femenino', 'Masculino' y la variable 'No_Especificia'

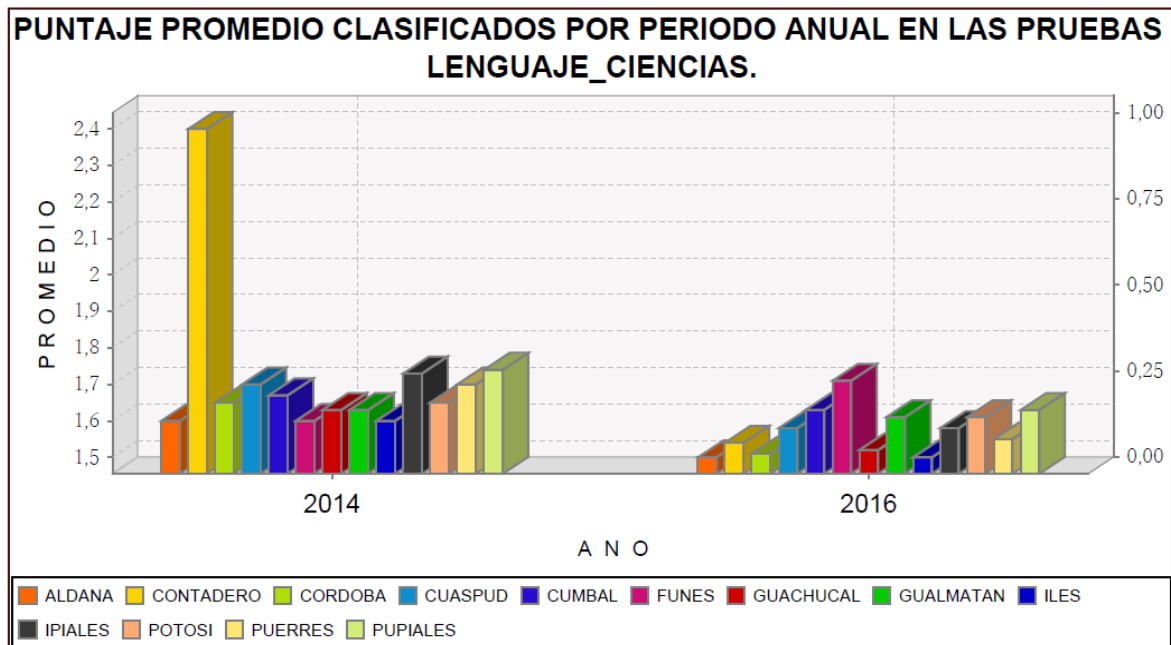
Figura 50. Porcentaje de estudiantes que presentaron la prueba de matemáticas y competencias, clasificada por género.



Fuente: Esta investigación.

¿Cuál es el puntaje promedio de los resultados obtenidos por año, en el área de lenguaje y ciencias? En la figura 51 se muestra los resultados utilizando Pentaho Report Designer.

Figura 51. Puntaje promedio de los resultados obtenidos por año en el área de lenguaje y ciencias.



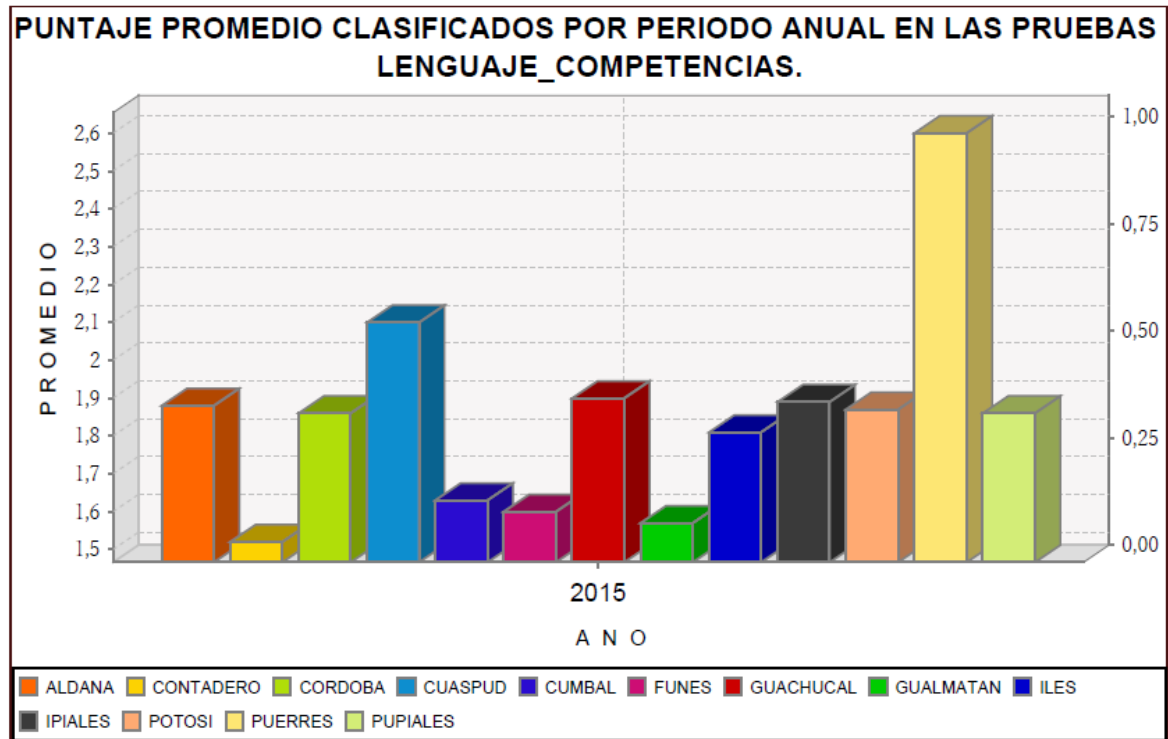
Fuente: Esta investigación.

De la figura 51 se deduce que la prueba de lenguaje y ciencias solo se presentó en los años 2014 y 2016. Siendo en el primer año el municipio de Contadero quien se destacó con un promedio alto con respecto a los demás municipios de la subregión de Obando, los cuales se encuentran bajo el promedio de 2.0 es el caso de IpiALES, Puerres y Pupiales mientras que Aldana, Funes e Iles se alejan considerablemente de un buen promedio.

En el año 2016 los promedios bajan con respecto a 2014, donde se observa que Contadero ya no presenta un buen promedio, al igual que IpiALES, Puerres y Pupiales, siendo esta una razón para revisar las políticas educativas de las Instituciones para mantener o aumentar el promedio que venían presentando.

¿Cuál es el puntaje promedio de los resultados obtenidos por año, en el área de lenguaje y competencias? En la figura 52 se muestra los resultados utilizando Pentaho Report Designer.

Figura 52. Puntaje promedio de los resultados obtenidos por año en el área de lenguaje y competencias.

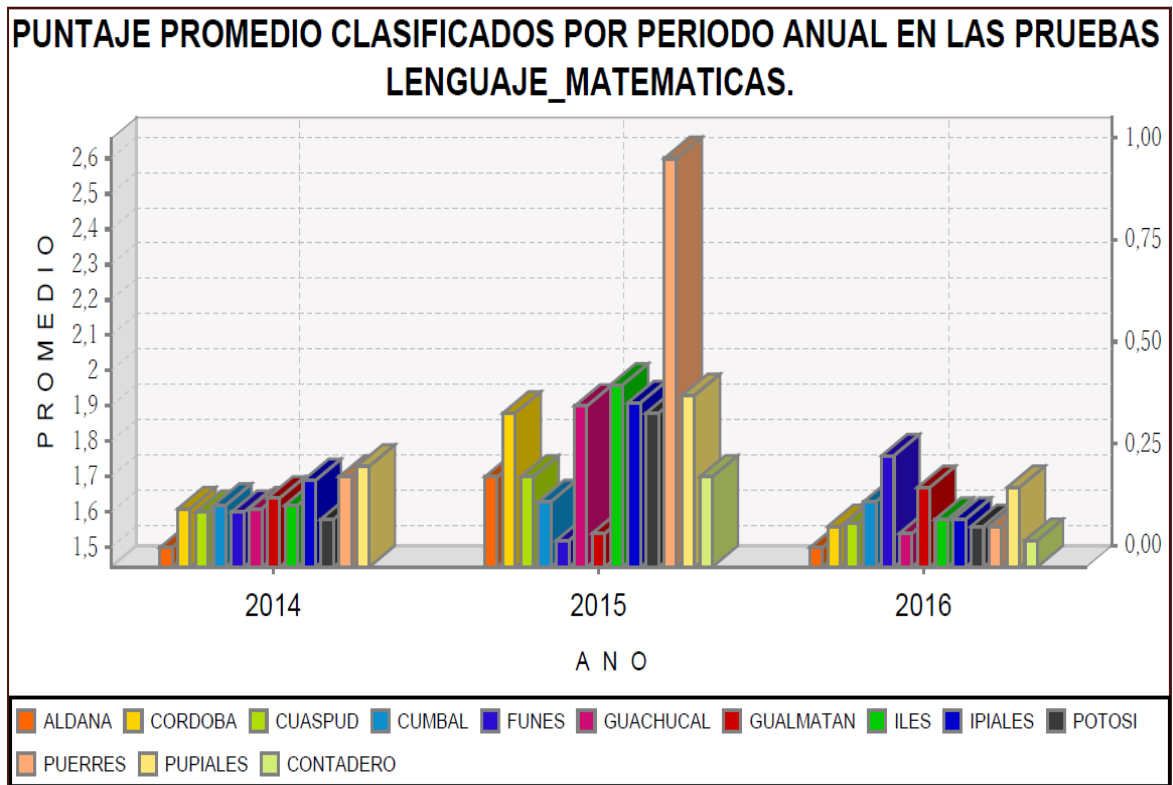


Fuente: Esta investigación.

De la figura 52 se observa que, lenguaje y matemáticas tan solo se presentó en un solo periodo, 2015. El comportamiento del promedio que se obtuvo en este año destaca a dos municipios, Puerres y Cuaspud que lograron superar el promedio de 2.0, mientras que Aldana, Córdoba, Guachucal, Ipiales, Potosí y Pupiales presentan un promedio aceptable con respecto a los municipios de Contadero, Cumbal, Funes y Gualmatan donde se debería buscar políticas de mejoramiento y así lograr obtener mejores resultados.

¿Cuál es el puntaje promedio de los resultados obtenidos por año, en el área de lenguaje y matemáticas? En la figura 53 se muestra los resultados utilizando Pentaho Report Designer.

Figura 53. Puntaje promedio de los resultados obtenidos por año en el área de lenguaje y matemáticas.



Fuente: Esta investigación.

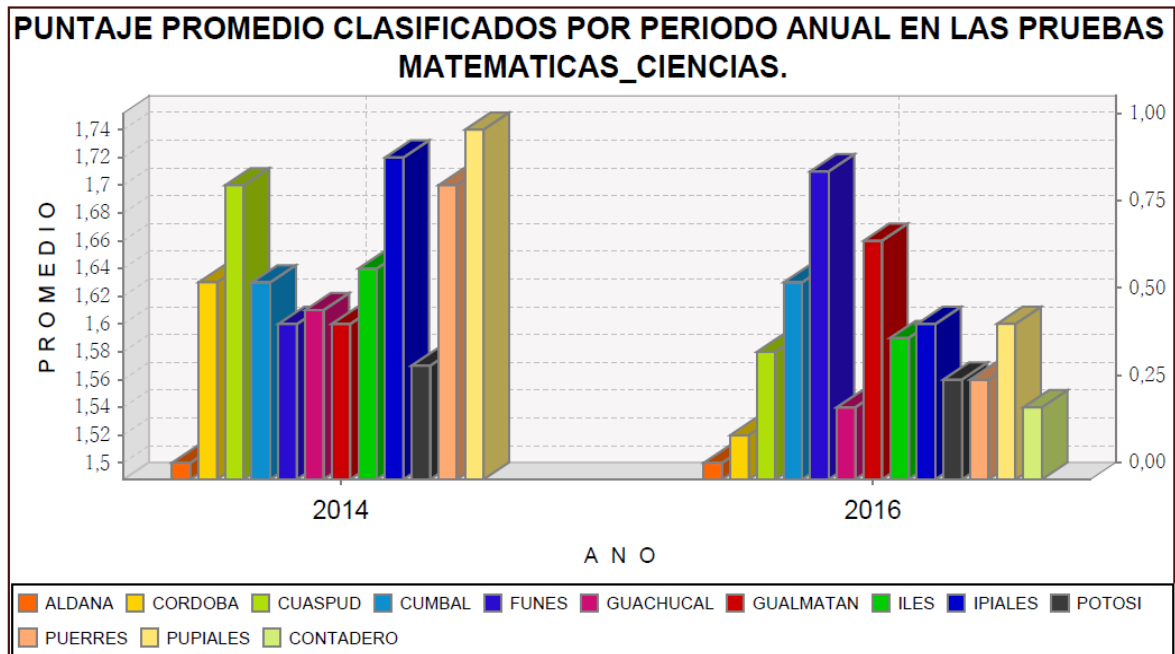
De acuerdo a la figura 53, lenguaje y matemáticas se presentó en los tres periodos En el año 2014, los promedios son considerablemente bajos en la mayoría de los municipios, mientras en Aldana y Potosí la situación es más crítica.

En 2015, se observa una mejoría de las Instituciones de la subregión de Obando, con respecto al anterior año, logrando así que municipios como Aldana y Potosí que anteriormente presentaban promedios bajos, esta vez logran aumentar su promedio, al igual que municipios como Córdoba, Guachucal, Iles, Ipiiales y Pupiales. En este año se destaca el municipio de Puerres con un promedio alto.

En 2016, las instituciones vuelven a bajar su rendimiento en las pruebas, los promedios son bajos e inaceptables en todos los municipios. En el caso particular de Aldana, Guachucal, Potosí y Contadero que en el anterior año lograron destacarse, esta vez se encuentran muy por debajo del buen promedio que traían.

¿Cuáles es el puntaje promedio de los resultados obtenidos por año, en el área de matemáticas y ciencias? En la figura 54 se muestra los resultados utilizando Pentaho Report Designer.

Figura 54. Puntaje promedio de los resultados obtenidos por año en el área de matemáticas y ciencias.



Fuente: Esta investigación.

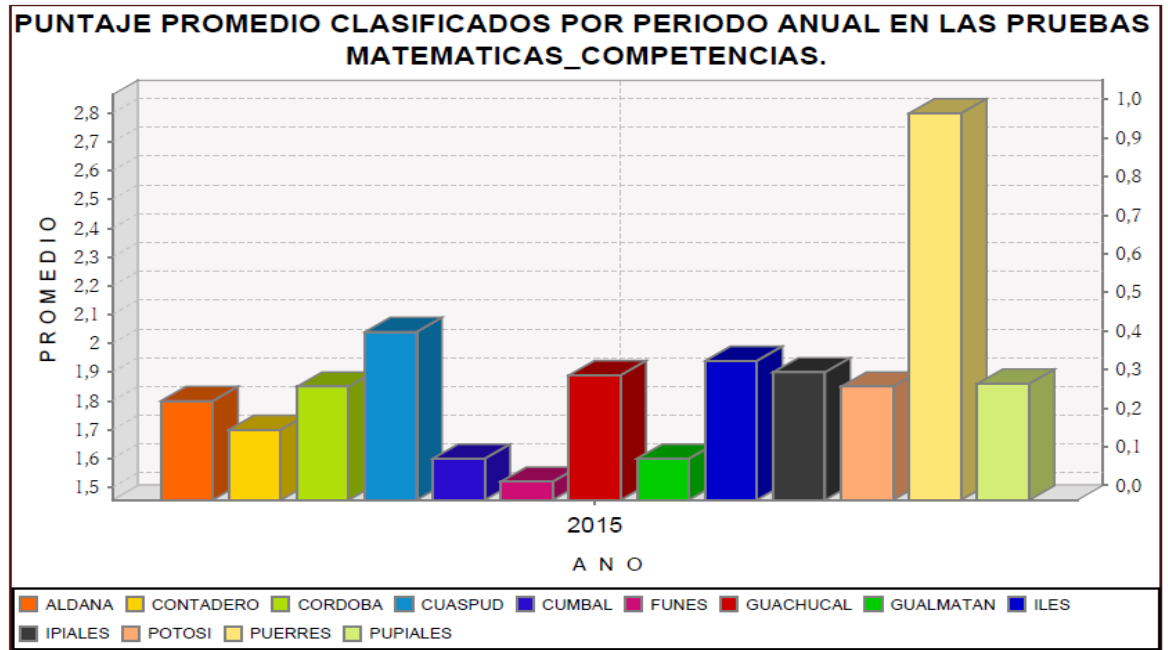
De la figura 54 se observa que matemáticas y ciencias no se presentó en el año 2015. Los promedios alcanzados por los municipios de la subregión de Obando son relativamente bajos en los años en que si se presentó la prueba ya que no logran estar por encima del promedio de 2.0

El comportamiento del municipio de Aldana en los dos años, es verdaderamente alarmante, ya que no presenta mejoría en el siguiente año y es uno de los municipios que tiene el promedio más bajo con respecto a los demás municipios.

Se puede destacar a los municipios de Funes y Guachucal que en el 2016 lograron superar el promedio que se obtuvo en 2014, Cumbal y Potosí lograron mantener el promedio en los dos periodos, mientras que municipios como Córdoba, Cuaspud, Guachucal, Iles, Puerres y Pupiales no lograron mantenerse y bajaron su promedio en 2016.

¿Cuál es el puntaje promedio de los resultados obtenidos por año, en el área de matemáticas y competencias? En la figura 55 se muestra los resultados utilizando Pentaho Report Designer.

Figura 55. Puntaje promedio de los resultados obtenidos por año en el área de matemáticas y competencias.



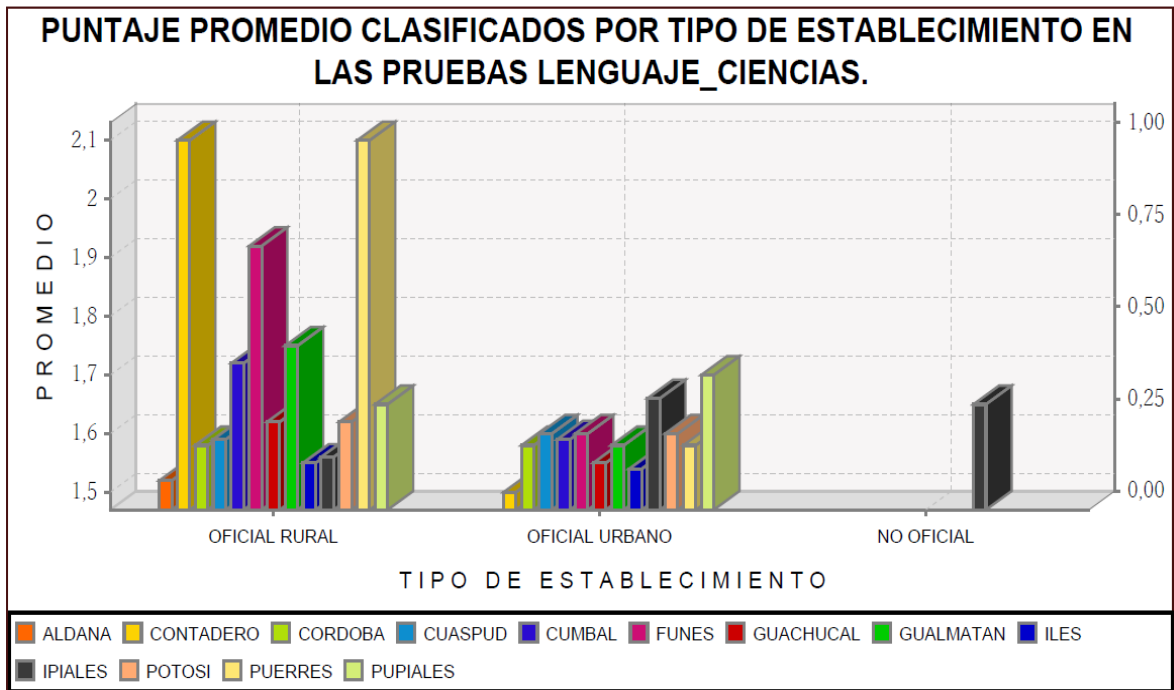
Fuente: Esta investigación.

De acuerdo a la figura 55 se observa que matemáticas y competencias solo se presentó en el año 2015, donde los promedios obtenidos por los municipios de la subregión de Obando son aceptables, salvo el caso de Cumbal, Funes y Gualmatan que presentan un promedio relativamente bajo, razón por la cual se deben tomar medidas de mejoramiento educativo para estar a la par del resto de municipios.

Nuevamente se destacan municipios como Puerres y Cuaspud, que logran superar el promedio de 2.0. Más abajo se encuentran Guachucal, Iles e Ipiales que con la ayuda de políticas educativas, se puede fortalecer el objetivo que es estar por encima del promedio de 2.0 al igual que el resto de municipios que presentan bajo rendimiento en la prueba.

¿Cuál es el puntaje promedio de los resultados en lenguaje y ciencias, clasificados por tipo de establecimiento? En la figura 56 se muestra los resultados utilizando Pentaho Report Designer.

Figura 56. Puntaje promedio de los resultados en lenguaje y ciencias, clasificados por tipo de establecimiento.



Fuente: Esta investigación.

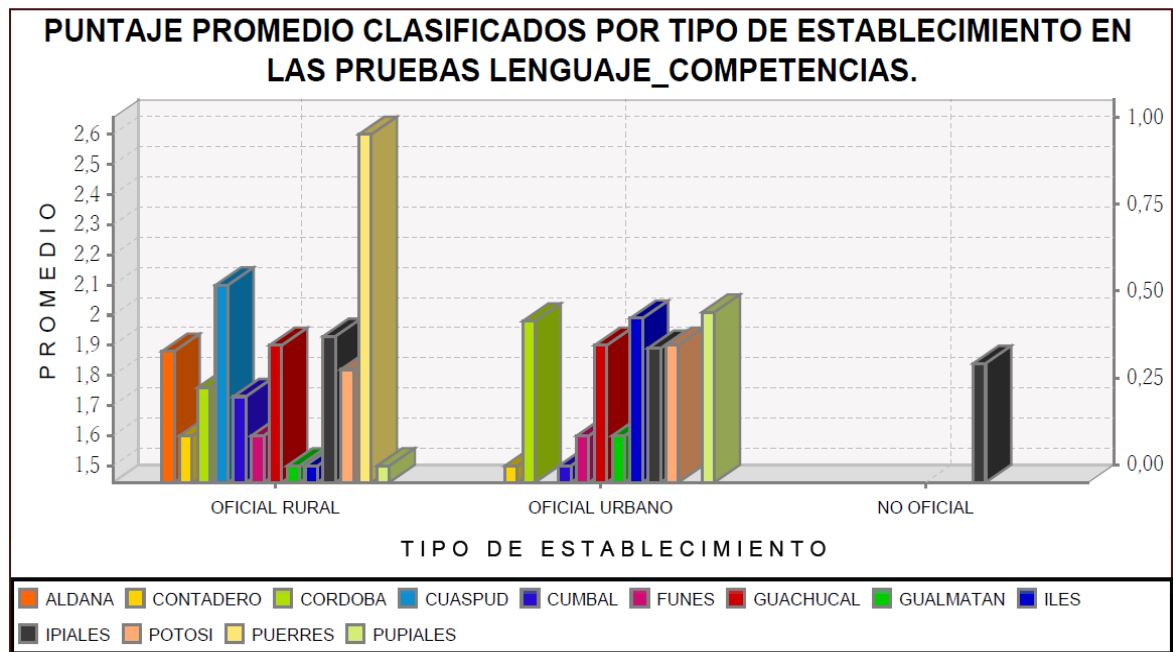
De la figura 56 observamos los promedios que obtuvieron los municipios de la subregión de Obando clasificados por tipo de establecimiento en lenguaje y ciencias. En Oficial Rural, el promedio alcanzado por los municipios es superior en relación a Oficial Urbano, se destaca Contadero y Puerres con promedios superiores a 2.0, pero también municipios como el caso de Aldana, Iles e Ipiales los promedios no son lo suficientemente acertados mirándose en la necesidad de tomar medidas de mejoramiento.

En Oficial Urbano sucede lo contrario con municipios como Contadero y Puerres que tenían promedios altos al igual que Funes aquí disminuyen drásticamente generando así alerta máxima de mejoramiento. Por otra parte, municipios como Iles y Potosí no presentan cambios, pero los promedios son bajos en general.

Por último, en el tipo No Oficial el municipio de Ipiales que al igual que en Oficial Rural y Oficial Urbano se mantiene el promedio que no supera el 2.0.

¿Cuál es el puntaje promedio de los resultados en lenguaje y competencias, clasificados por tipo de establecimiento? En la figura 57 se muestra los resultados utilizando Pentaho Report Designer.

Figura 57. Puntaje promedio de los resultados en lenguaje y competencias, clasificados por tipo de establecimiento.



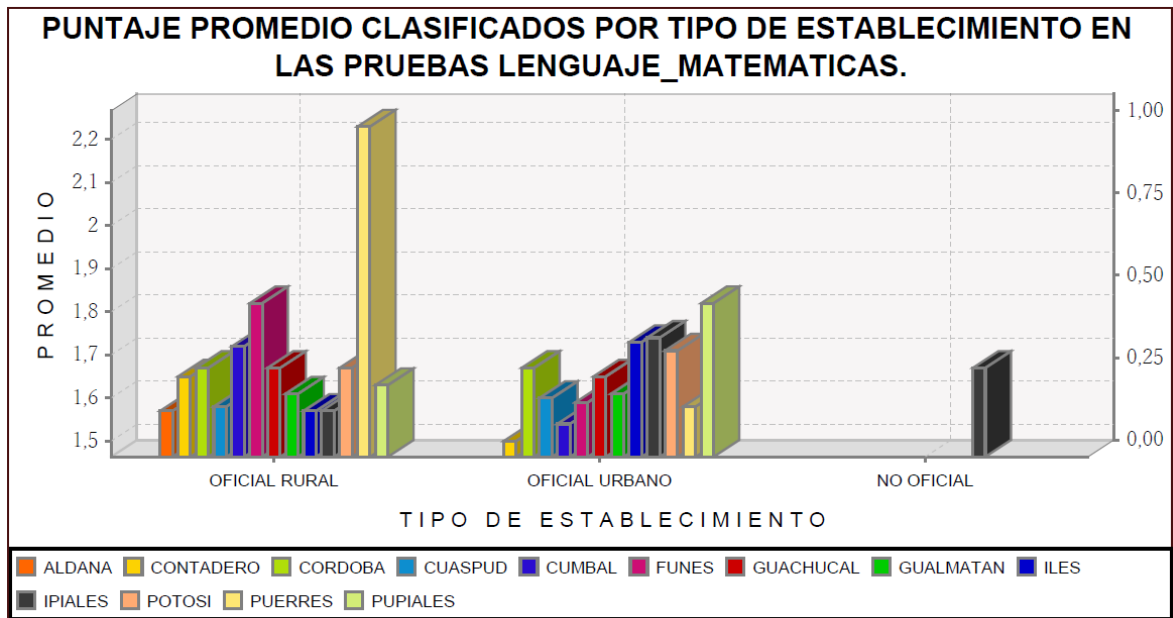
Fuente: Esta investigación.

De la figura 57 observamos los promedios que obtuvieron los municipios de la subregión de Obando clasificados por tipo de establecimiento en lenguaje y competencias. En Oficial Rural se destaca ampliamente el municipio de Puerres frente al resto de municipios, pero en general aumento el número de municipios con un promedio aceptable como el caso de Aldana, Cuaspud, Guachucal, Ipiales y Potosí que obtuvieron un promedio por debajo de 2.0 con mínima diferencia, salvo el caso de Gualmatán, Iles y Pupiales cuyos promedios resultan alarmantes frente al resto de la subregión. En Oficial Urbano el promedio de los municipios resulta parejo y de igual forma aceptable como el caso de Córdoba, Funes, Iles y Pupiales, este último municipio presenta una gran diferencia frente al Oficial Rural donde se observa una gran diferencia de promedio llegando a 2.0.

En No Oficial el municipio de Ipiales presenta un promedio aceptable con respecto a los municipios de Oficial Rural y Oficial Urbano.

¿Cuál es el puntaje promedio de los resultados en lenguaje y matemáticas, clasificados por tipo de establecimiento? En la figura 58 se muestra los resultados utilizando Pentaho Report Designer.

Figura 58. Puntaje promedio de los resultados en lenguaje y matemáticas, clasificados por tipo de establecimiento.



Fuente: Esta investigación.

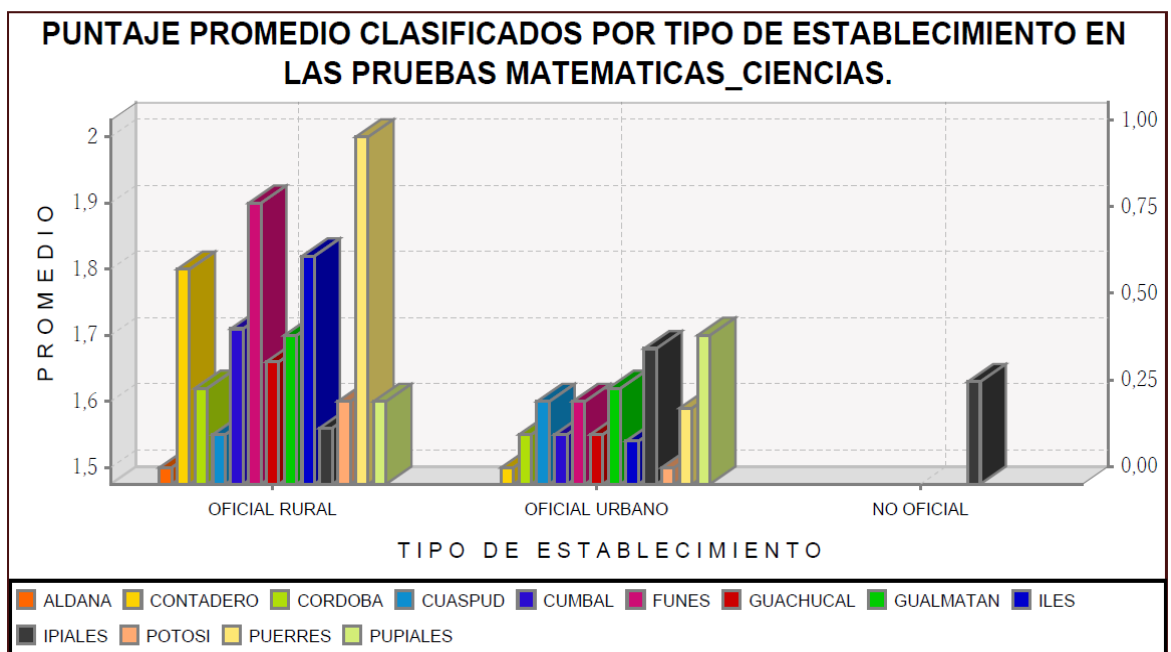
De acuerdo a la figura 58 donde se observa los promedios que obtuvieron los municipios de la subregión de Obando clasificados por tipo de establecimiento en lenguaje y matemáticas, se concluye que en general los promedios son parejos tanto en Oficial Rural como en Oficial Urbano, pero relativamente bajos. Mientras que en el primero se destaca nuevamente Puerres con un promedio superior a 2.0, en Oficial Urbano el promedio baja drásticamente, otro caso similar es el de los municipios de Contadero y Funes donde su promedio desciende en Oficial Urbano.

Por otro lado, sucede todo lo contrario con los municipios de Iles, Ipiales y Potosí que presentan promedios bajos en Oficial Rural, mientras que en Oficial Urbano logran aumentar su promedio.

Finalmente, en No Oficial el promedio del municipio de Ipiales se mantiene estable con respecto a los otros dos tipos de establecimientos es decir el promedio no es muy alto ni muy bajo.

¿Cuál es el puntaje promedio de los resultados en matemáticas y ciencias, clasificados por tipo de establecimiento? En la figura 59 se muestra los resultados utilizando Pentaho Report Designer.

Figura 59. Puntaje promedio de los resultados en matemáticas y ciencias, clasificados por tipo de establecimiento.



Fuente: Esta investigación.

De acuerdo a la figura 59 donde se observa los promedios que obtuvieron los municipios de la subregión de Obando clasificados por tipo de establecimiento en matemáticas y ciencias, se concluye que las Instituciones Educativas de Oficial Rural presentan mayores promedios que Oficial Urbano, pero los promedios no son tan favorables para la subregión, ya que la mayoría no logran alcanzar el promedio de 2.0 a excepción del municipio de Puerres.

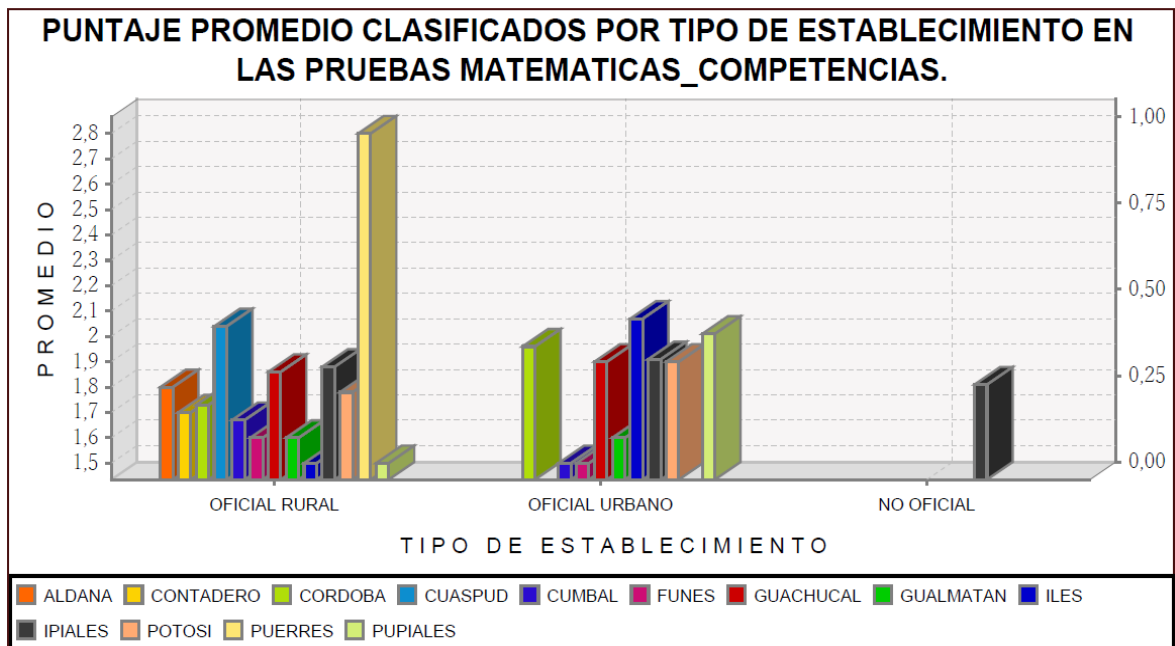
En Oficial Rural municipios como Contadero, Córdoba, Funes, Guachucal, Iles y Potosí presentaban promedios razonables, pero disminuyen considerablemente en Oficial Urbano, lo cual significa que habría que tomar medidas para entender ese drástico cambio de bajo rendimiento. Caso contrario ocurre con los municipios de

Cuaspud, Ipiales y Pupiales donde se logra mejorar el promedio en Oficial Urbano con respecto a Oficial Rural.

Por último, el promedio de Ipiales en No Oficial es mejor que el obtenido en Oficial Rural pero no en Oficial Urbano, como conclusión general el promedio en matemáticas y ciencias es bajo.

¿Cuál es el puntaje promedio de los resultados en matemáticas y competencias, clasificados por tipo de establecimiento? En la figura 60 se muestra los resultados utilizando Pentaho Report Designer.

Figura 60. Puntaje promedio de los resultados en matemáticas y competencias, clasificados por tipo de establecimiento.



Fuente: Esta investigación.

A partir de la figura 60 se observa los promedios que obtuvieron los municipios de la subregión de Obando clasificados por tipo de establecimiento en matemáticas y competencias, se concluye que, en Oficial Urbano los municipios obtienen un promedio mayor al conseguido por los de tipo Oficial Rural

En Oficial Rural se destaca los municipios de Puerres y Cuaspud con promedios que superan el 2.0, también se hace énfasis al comportamiento que presentan Aldana, Guachucal e Ipiales donde la diferencia es mínima para alcanzar un

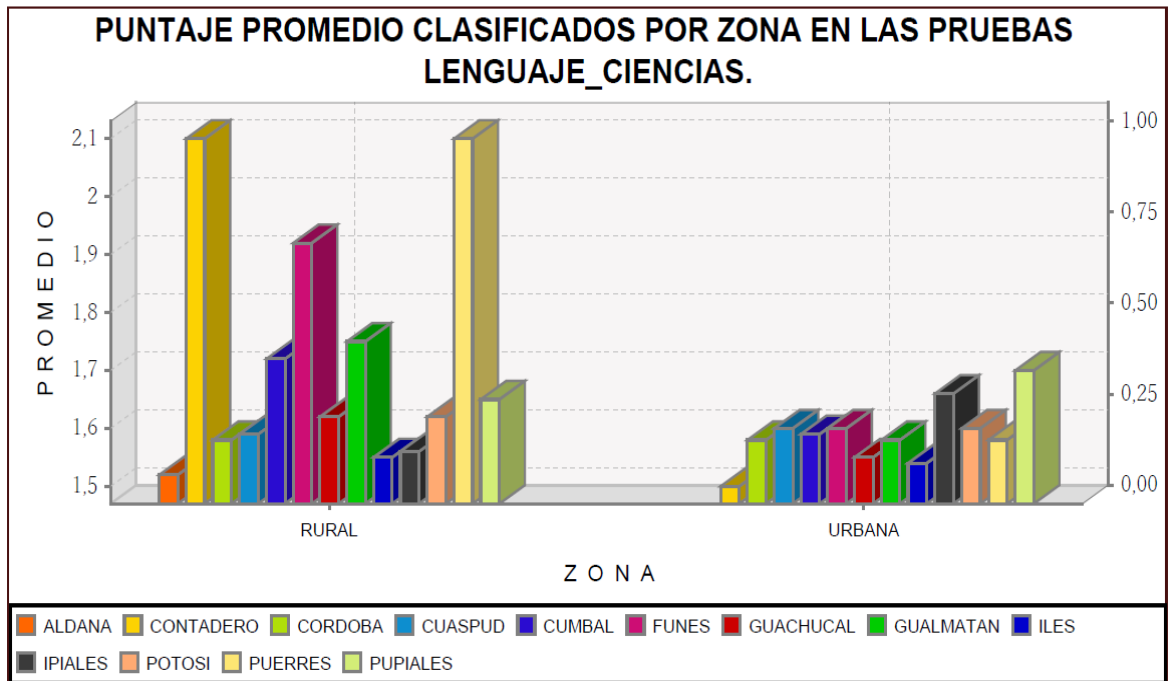
promedio aceptable, cosa que no ocurre con Iles y Pupiales los cuales representan alerta máxima de mejoramiento pues sus promedios son bajos.

En Oficial Urbano, los municipios de Guachucal, Ipiales, y Potosí logran mejorar sus promedios con respecto a Oficial Rural y alcanzan un promedio de 2.0 siendo estos puntajes aceptables, mientras que Córdoba, Iles y Potosí sobrepasan este promedio, contrario a lo que no sucede con Cumbal y Funes que se encuentran en alerta máxima obligando así a las secretarías de educación a tomar medidas lo más pronto posible.

Por último, en No Oficial el promedio que consigue el municipio de Ipiales se acerca al promedio de 2.0 con una mínima diferencia, donde bastarían ciertas medidas de mejoramiento para estar en un promedio aceptable.

¿Cuál es el puntaje promedio de los resultados en lenguaje y ciencias, clasificados por zona? En la figura 61 se muestra los resultados utilizando Pentaho Report Designer.

Figura 61. Puntaje promedio de los resultados en lenguaje y ciencias, clasificados por zona.



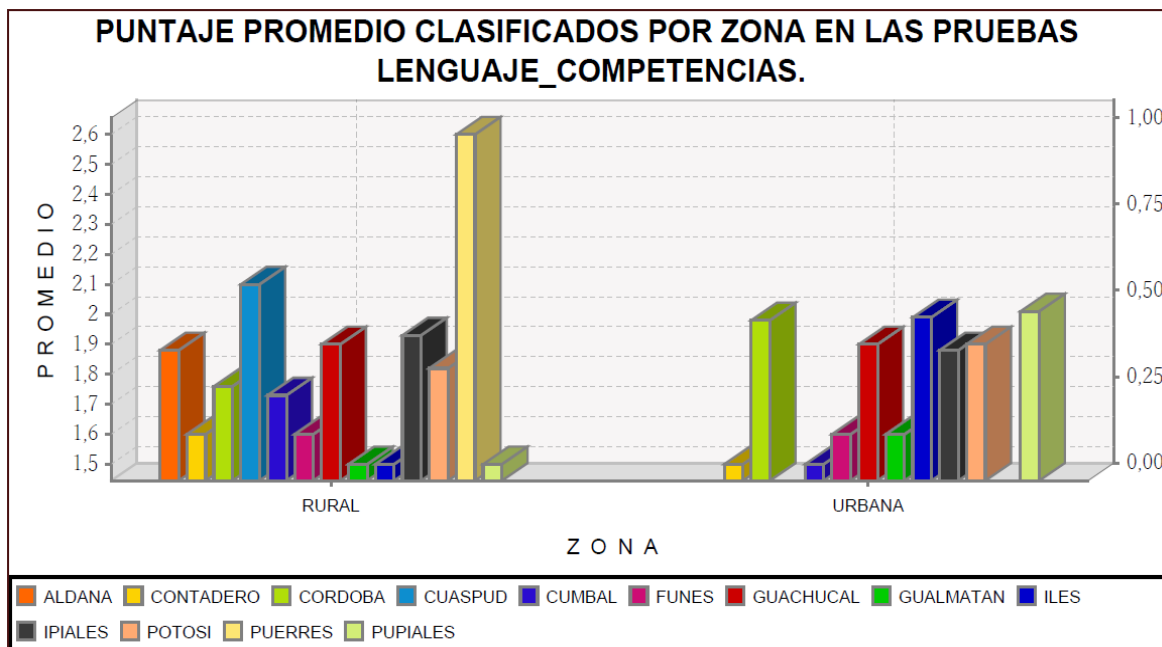
Fuente: Esta investigación.

A partir de la figura 61 se observa los promedios que obtuvieron los municipios de la subregión de Obando clasificados por zona, en lenguaje y ciencias, donde claramente se ve que los promedios alcanzados por las Instituciones Educativas de la zona rural son mayores con relación a los de zona urbana.

En la zona rural se destacan Contadero y Puerres con promedios que superan el 2.0, mientras que Funes se aleja con una mínima diferencia, en el caso de Aldana, Iles e Ipiiales son los municipios con los promedios más bajos respecto a los demás municipios. Mientras que en la zona Urbana los municipios que se destacaron en lo rural, bajaron drásticamente su promedio, es el caso de Contadero que llegó a tener el promedio más bajo, de igual forma lo hicieron Cumbal, Funes, Guachucal, Gualmatan y Puerres, lo cual se convierte en una razón para analizar este tipo de comportamientos que se presentan en la zona rural y urbana.

¿Cuál es el puntaje promedio de los resultados en lenguaje y competencias, clasificados por zona? En la figura 62 se muestra los resultados utilizando Pentaho Report Designer.

Figura 62. Puntaje promedio de los resultados en lenguaje y competencias, clasificados por zona.



Fuente: Esta investigación.

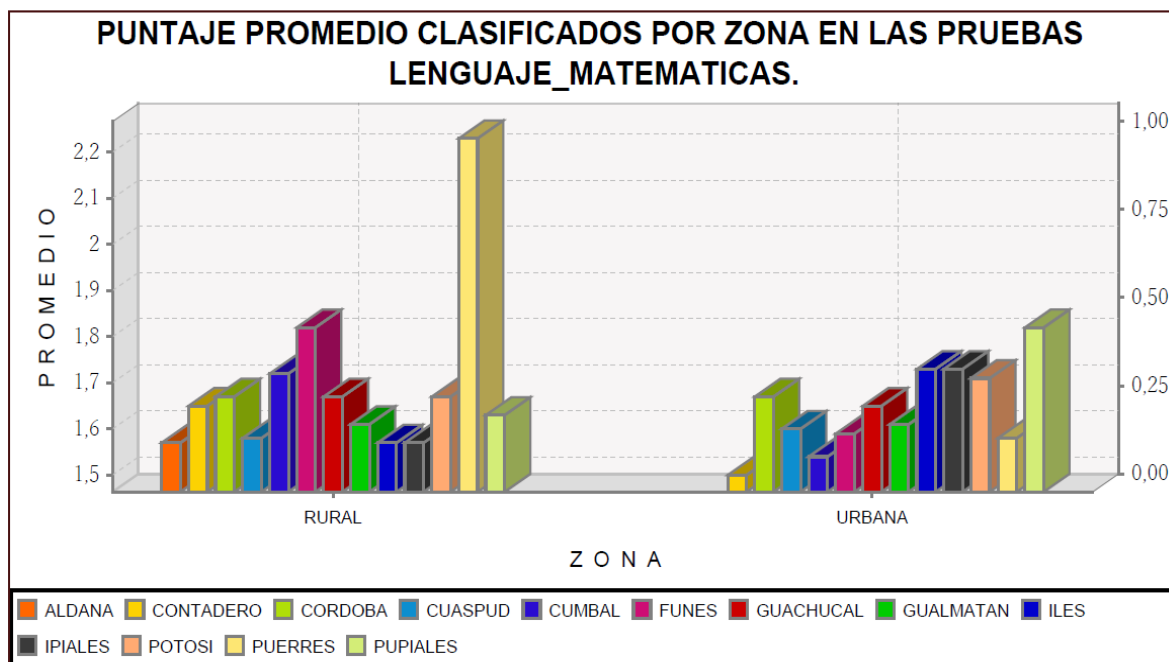
A partir de la figura 62 se observa los promedios que obtuvieron los municipios de la subregión de Obando clasificados por zona, en lenguaje y competencias donde el comportamiento en la zona rural como urbana es casi que igual.

En la zona rural los municipios de Puerres y Cuaspud logran obtener promedios que superan el 2.0, mientras que Aldana, Guachucal e Ipiales se encuentran por debajo de este promedio con una mínima diferencia, el caso alarmante lo representan los municipios de Gualmatan, Iles y Pupiales donde se ve la necesidad de tomar medidas para mejorar la calidad educativa.

En la zona urbana los municipios de Córdoba, Guachucal, Iles y Pupiales se destacan con promedios aceptables, y mayores a los de la zona rural, mientras que Contadero y Cumbal se ven en la necesidad de replantear las políticas educativas para estar a la par del resto de municipios de la subregión.

¿Cuál es el puntaje promedio de los resultados en lenguaje y matemáticas, clasificados por zona? En la figura 63 se muestra los resultados utilizando Pentaho Report Designer.

Figura 63. Puntaje promedio de los resultados en lenguaje y matemáticas, clasificados por zona.



Fuente: Esta investigación.

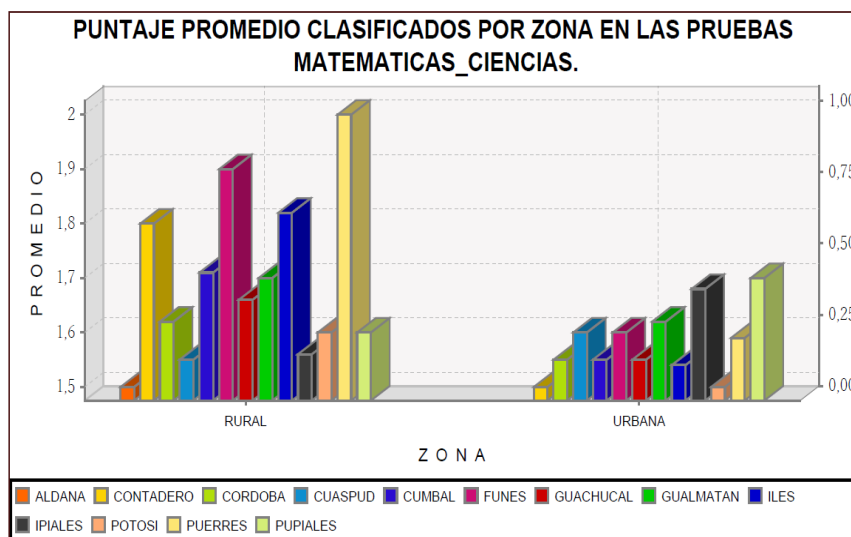
A partir de la figura 63 se observa los promedios que obtuvieron los municipios de la subregión de Obando clasificados por zona, en lenguaje y matemáticas, donde es posible afirmar que los promedios en zona rural como urbana son bajos.

En zona rural, nuevamente se destaca el municipio de Puerres con un promedio relativamente mayor frente al resto de municipios que no logran acercarse a un promedio aceptable de 2.0. Cómo Cumbal y Funes, mientras que Aldana, Cuaspuj, Iles e Ipiales siguen presentando promedios bajos.

En zona urbana, como se decía anteriormente, los promedios a nivel general son bajos, donde no existe ningún municipio que logre estar con un promedio mayor o igual a 2.0. El caso de Puerres en particular presenta un comportamiento especial, si bien en zona rural logra destacarse con un promedio aceptable, no se explica cómo en zona urbana el promedio desciende drásticamente, tal es el caso de municipios como Contadero, Cumbal, y Funes que presentan un comportamiento similar. Caso contrario sucede con Iles, Ipiales, Potosí y Pupiales donde logran aumentar su promedio con respecto al que obtuvieron las instituciones de zona rural.

¿Cuál es el puntaje promedio de los resultados en matemáticas y ciencias, clasificados por zona? En la figura 64 se muestra los resultados utilizando Pentaho Report Designer.

Figura 64. Puntaje promedio de los resultados en matemáticas y ciencias, clasificados por zona.



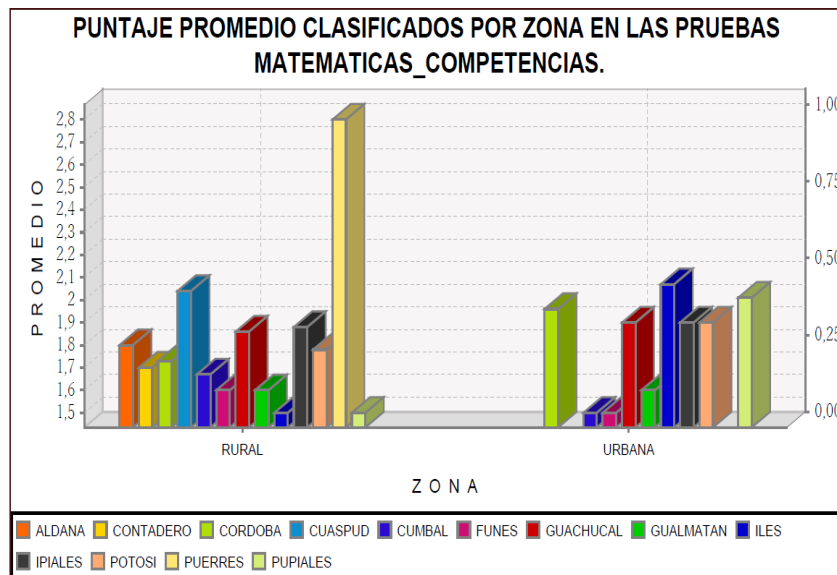
Fuente: Esta investigación.

A partir de la figura 64 se observa los promedios que obtuvieron los municipios de la subregión de Obando clasificados por zona, en matemáticas y ciencias, donde los promedios en zona rural claramente son mayores con respecto a zona urbana. En la zona rural las instituciones del municipio de Puerres se han venido destacando en las diferentes áreas del conocimiento, con promedios mayores o iguales a 2.0, acontecimiento que no sucede en el resto de municipios de la subregión. Mientras que municipios como Contadero, Funes e Iles se encuentran por debajo del promedio de 2.0 con una mínima diferencia. Pero en el caso opuesto se encuentran Aldana, Cuaspud e Ipiales con bajos promedios.

En la zona Urbana, se presenta un comportamiento anormal, donde las instituciones educativas están descendiendo su promedio conseguido en zona rural, es el caso de Contadero, Córdoba, Cumbal, Funes, Gualmatan, Iles, Potosí y Puerres resultando así una razón para que las secretarías de educación busquen tomar alternativas de mejoramiento para lograr mantener o ascender el promedio que se está consiguiendo en zona rural, para que no siga ocurriendo acontecimientos de este tipo.

¿Cuál es el puntaje promedio de los resultados en matemáticas y competencias, clasificados por zona? En la figura 65 se muestra los resultados utilizando Pentaho Report Designer.

Figura 65. Puntaje promedio de los resultados en matemáticas y competencias, clasificados por zona.



Fuente: Esta investigación.

A partir de la figura 65 se observa los promedios que obtuvieron los municipios de la subregión de Obando clasificados por zona, en matemáticas y competencias, donde las Instituciones Educativas de la zona urbana aventajan a las de zona rural con una mínima diferencia en sus promedios.

En la zona rural, Puerres se destaca con un promedio significativamente alto con respecto a los demás municipios y al igual que Cuaspud logran superar el promedio de 2.0, en el caso de Aldana, Guachucal e Ipiales se alejan con una mínima diferencia del promedio que lograron superar los anteriores municipios.

Por otra parte, los promedios obtenidos por Iles y Pupiales que resultan ser los más bajos, es una medida para tomar decisiones de mejoramiento y dado el caso replantear las políticas educativas que se manejan internamente para lograr obtener promedios acordes al promedio ideal.

En la zona urbana, si bien, el comportamiento de promedio es más parejo y aceptable con municipios como Córdoba, Guachucal, Iles, Ipiales, Potosí y Pupiales también se presentan casos similares a los de zona rural, siendo Cumbal y Funes los municipios en alerta máxima de mejoramiento.

3. CONCLUSIONES Y RECOMENDACIONES

El objetivo de este proyecto de crear un mercado de datos para el almacenamiento histórico de los resultados de los estudiantes de grado 9º en las pruebas saber 9º, como una herramienta para soportar las decisiones de los entes gubernamentales y directivos educativos de la subregión de Obando se cumplió en su totalidad.

Como resultado de ello la subregión de Obando cuenta con un mercado de datos con los resultados de las pruebas saber 9º de los estudiantes de grado 9º de esta subregión, que le permitirá conjuntamente con la herramienta Pentaho realizar un análisis por múltiples dimensiones o variables y visualizar sus resultados de manera gráfica, facilitando la toma de decisiones.

En este proyecto, la etapa que más consumió recursos y tiempo fue la de extracción, transformación y carga de los datos de las bases de datos del ICFES al mercado de datos, por la mala calidad de los datos que implicó su limpieza y transformación.

Es necesario documentar cada una de las fases del proyecto y generar los metadatos correspondientes a fin de poder replicar el proceso en trabajos futuros de la misma naturaleza

Sería de mucho beneficio que todas las Instituciones Educativas del país, dispongan de su propio mercado de datos, con información histórica, limpia y transformada de los resultados de las pruebas saber, con el objetivo de conocer las fortalezas y debilidades que presenta la calidad educativa del país, permitiendo a las secretarías de educación, establecimientos educativos y al Ministerio de Educación Nacional, tomar medidas de mejoramiento y así lograr el objetivo primordial de la educación nacional, que Colombia sea el país más educado en América en el 2025.

El contar con un mercado de datos que almacena la información histórica de los estudiantes de grado 9º que presentaron las pruebas Saber 9º en los años 2014,

2015 y 2016 de las Instituciones Educativas de la subregión de Obando, se constituye en un hecho muy importante para estas instituciones, por la calidad de información que se puede obtener. Ya es función de cada municipio actualizar este mercado con los datos de los subsiguientes años. Por esta razón se plantea como trabajo futuro la construcción de un aplicativo web que se conecte al mercado de datos Saber 9^o construido y que permita a los municipios de la subregión de Obando actualizar la información de los resultados de las pruebas saber 9 en los años por venir.

BIBLIOGRAFÍA

ICFES (2011) Lineamientos generales Saber 2009 5° y 9°. Instituto Colombiano para la Evaluación de la Educación (ICFES). ISBN de la versión electrónica: 978-958-11-0489-5. Bogotá, marzo de 2009.

ICFES (2011) SABER 5° y 9° 2009 Síntesis de resultados de factores asociados. Instituto Colombiano para la Evaluación de la Educación (ICFES). ISBN de la versión electrónica: 978-958-11-0574-8. Bogotá, Colombia.

ICFES (2011). Informe técnico Saber 5° y 9° 2009. Instituto Colombiano para la Evaluación de la Educación (ICFES). ISBN de la versión electrónica: 978-958-11-0578-6. Bogotá, Colombia.

ICFES (2011). SABER 5° y 9° 2009 Síntesis de resultados de factores asociados. Instituto Colombiano para la Evaluación de la Educación (ICFES). ISBN de la versión electrónica: 978-958-11-0574-8. Bogotá, Colombia.

ICFES (2011). Marco de factores Asociados Saber 3°, 5° y 9° 2016. Instituto Colombiano para la Evaluación de la Educación (ICFES). ISBN de la versión digital: En trámite. Bogotá, Colombia.

ICFES (2013). Alineación del examen SABER 11°. Sistema Nacional de Evaluación Estandarizada de la Educación, Instituto Colombiano para la Evaluación de la Educación (ICFES). Bogotá, Colombia.

ICFES (2014). Alineación del examen SABER 11° Lineamientos generales 2014-2 Sistema Nacional de Evaluación Estandarizada de la Educación, Instituto Colombiano para la Evaluación de la Educación (ICFES). ISBN: 9789581106301. Bogotá, Colombia.

Espiñera, Sheldon y Asociados. La Inteligencia de Negocios (Business Intelligence). Boletín e Asesoría Gerencial. 2008. [Consultado el 30 de Setiembre 2011] <<http://www.pwc.com/ve/es/asesoria-gerencial/boletin/assets/boletin-advisory-edicion-10-2008.pdf>>

Lluís, Josep. (2008). *Business Intelligence: Competir con Información*. Barcelona-España: Editorial Dataprix.

Dresner, H. (2014). *Glosario términos: BI*. Gartner. Recuperado de <http://www.gartner.com/it-glossary/business-intelligence-bi..>

Sperley, Eric. "The Enterprise Data Warehouse: Planning, Building, and Implementation". Prentice Hall. 1999.

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.

Kimball, R, & Ross, M, *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling* (Third Edition), Indianapolis, Wiley, 2013.

Riquelme, M, ¿Qué es un datamart? | Web y Empresas. [online] Available at: <https://www.webyempresas.com/que-es-un-data-mart/>

Chinchilla Arley, Ricardo; Vargas Murillo, José Luis. (2002) Prototipo de Mercado de Datos para la División de Control y Fiscalización de la Dirección General de Aduanas

Rivadera, 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.

Mundy & Thornthwaite, *The Microsoft Data Warehouse Toolkit— With SQL Server 2005 and the Microsoft Business Intelligence Toolset*, Indianapolis, Wiley, 2006.

<https://www.semana.com/educacion/articulo/en-que-consisten-las-pruebas-saber/495713>

Pagina web Gobernación de Nariño enlace <http://xn--nario-rta.gov.co/2012-2015/index.php/ex-providencia-de-obando>

Proyección de Población municipales por área. Disponible en https://www.dane.gov.co/files/investigaciones/poblacion/proyepobla06_20/ProyeccionMunicipios2005_2020.xls.

ANEXOS

ANEXO 1. INTEGRACIÓN DE DATOS

- 1) Descargar de la página ICFES los repositorios de las pruebas saber 9.
- 2) Se creó la base de datos **PruebaSaber9** en el SGBD PostgreSQL.
- 3) En la base de datos **PruebaSaber9** se creó las tablas
 - comp_ent_terr(2014,2015,2016).
 - comp_estab(2014,2015,2016).
 - comp_municipio(2014,2015,2016).
 - comp_sede(2014,2015,2016).
 - Plausibles(2014,2015,2016).
- 4) Se realizó la carga de los archivos planos al gestor SGBD PostgreSQL según las tablas anteriores.
- 5) Adicionar en las tablas plausibles2014, plausibles2015, plausibles2016 el atributo ano de tipo text:
alter table plausibles2014 add ano text;
alter table plausibles2015 add ano text;
alter table plausibles2016 add ano text;
- 6) Actualizar la tabla plausibles2014, plausibles2015, plausibles2016 el atributo ano con el valor del año en que se realizó la prueba:
Update plausibles2014 set ano=2014;
Update plausibles2015 set ano=2015;
Update plausibles2016 set ano=2016;
- 7) Se creó las tablas
 - comp_ent_terr2014_2016 a partir de comp_ent_terr2016
create table comp_ent_terr2014_2016 as select * from comp_ent_terr2016;
 - comp_estab2014_2016 a partir de comp_estab2016.
create table comp_estab2014_2016 as select * from comp_estab2016;

- comp_municipio2014_2016 a partir de comp_municipio2016
create table comp_municipio2014_2016 as select * from comp_municipio2016;
- comp_sede2014_2016 a partir de comp_sede2016.
create table comp_sede2014_2016 as select * from comp_sede2016;
- plausibles_2014_2016 a partir de las tablas mencionadas anteriormente.
create table plausibles_2014_2016 as select * from plausibles2016;

8) Visualizar los esquemas de las tablas con el fin de obtener los atributos comunes:

ENTE TERRITORIAL

➤ *Cambiar comp_ent2014_terr2015, comp_estab2016*

```
select column_name from information_schema.columns where
table_name='comp_ent_terr2014_2016';
```

```
select column_name from information_schema.columns where
table_name='comp_ent_terr2014';
```

```
select column_name from information_schema.columns where
table_name='comp_ent_terr2015';
```

```
select column_name from information_schema.columns where
table_name='comp_ent_terr2016';
```

```
select column_name from information_schema.columns where
table_name='comp_ent_terr2014_2016'
intersect
select column_name from information_schema.columns where
table_name='comp_ent_terr2016';
```

ENTE TERRITORIAL

➤ *Cambiar comp_estab2015, comp_estab2016*

```
select column_name from information_schema.columns where  
table_name='comp_estab2014_2016';
```

```
select column_name from information_schema.columns where  
table_name='comp_estab2014';
```

```
select column_name from information_schema.columns where  
table_name='comp_estab2014';
```

```
select column_name from information_schema.columns where  
table_name='comp_estab2014';
```

```
select column_name from information_schema.columns where  
table_name='comp_estab2014_2016'  
intersect  
select column_name from information_schema.columns where  
table_name='comp_estab2014';
```

MUNICIPIO

➤ *cambiar comp_municipio2014, comp_municipio2015*

```
select column_name from information_schema.columns where  
table_name='comp_municipio2014_2016';
```

```
select column_name from information_schema.columns where  
table_name='comp_municipio2014';
```

```
select column_name from information_schema.columns where  
table_name='comp_municipio2015';
```

```
select column_name from information_schema.columns where  
table_name='comp_municipio2016';
```

```
select column_name from information_schema.columns where  
table_name='comp_municipio2014_2016'  
intersect
```

```
select column_name from information_schema.columns where
table_name='comp_municipio2016';
```

SEDES

➤ *Cambiar comp_sede2014, comp_sede2015*

```
select column_name from information_schema.columns where
table_name='comp_sede2014_2016';
```

```
select column_name from information_schema.columns where
table_name='comp_sede2014';
select column_name from information_schema.columns where
table_name='comp_sede2015';
```

```
select column_name from information_schema.columns where
table_name='comp_sede2016';
```

```
select column_name from information_schema.columns where
table_name='comp_sede2014_2016'
intersect
select column_name from information_schema.columns where
table_name='comp_sede2016';
```

PLAUSIBLES

➤ *Cambiar plausibles2014, plausibles2015.*

```
select column_name from information_schema.columns where
table_name='plausibles_2014_2016';
```

```
select column_name from information_schema.columns where
table_name='plausibles2014';
```

```
select column_name from information_schema.columns where
table_name='plausibles2015';
```

```
select column_name from information_schema.columns where
table_name='plausibles2016';
```

```

select column_name from information_schema.columns where
table_name='plausibles_2014_2016'
intersect
select column_name from information_schema.columns where
table_name='plausibles2016';

```

9) Insertar los datos de las tablas

- comp_ent_terr(2014,2015) a la tabla comp_ent_terr2014_2016.

```

INSERT INTO comp_ent_terr2014_2016
(id_ente,nombre,munexclu,tipo)
select * from comp_ent_terr2015 where
id_ente not in (select id_ente from comp_ent_terr2014_2016);

```

En la tabla comp_ent_terr2014 no existen datos diferentes a comp_ent_terr2016.

- Se cambió en el campo munexclu los nulos por sin dato.

```

update comp_ent_terr2014_2016 set munexclu='SIN_DATO' WHERE
munexclu=null;

```

- Se convirtió los campo nombre,munexclu de minúsculas a mayúsculas.

```

update comp_ent_terr2014_2016 set
nombre=upper(translate(nombre,'áéíóúÁÉÍÓÚäëïöüÄËÏÖÜÑ','aeiouAEIOUa
eiouAEIOUN'));

```

```

update comp_ent_terr2014_2016 set munexclu
=upper(translate(munexclu,'áéíóúÁÉÍÓÚäëïöüÄËÏÖÜÑ','aeiouAEIOUaeiouA
EIOUN'));

```

- comp_estab(2014,2015) a la tabla comp_estab2014_2016.

```

INSERT INTO comp_estab2014_2016
(cod_dane,id_municipio,id_ente,nombre,zona,sector,
tipo_estab,calendario,nivel_socio)
select * from comp_estab2014 where

```

```
cod_dane not in (select cod_dane from comp_estab2014_2016);
```

```
INSERT INTO comp_estab2014_2016  
(cod_dane,id_municipio,id_ente,nombre,zona,sector,  
tipo_estab,calendario,nivel_socio)  
select * from comp_estab2015 where  
cod_dane not in (select cod_dane from comp_estab2014_2016);
```

- comp_municipio(2014) a la tabla comp_municipio2014_2016.

```
INSERT INTO comp_municipio2014_2016  
(id_ente,nombre,munexclu,tipo)  
select * from comp_municipio2014 where  
id_ente not in (select id_ente from comp_municipio2014_2016);
```

En la tabla comp_municipio2015 no existen datos diferentes a comp_municipio2014_2016.

- El campo munexclu se elimino porq todos están vacíos.
alter table comp_municipio2014_2016 drop munexclu;
- Se convirtió el campo nombre de minúsculas a mayúsculas.
update comp_municipio2014_2016 set
nombre=upper(translate(nombre,'áéíóúÁÉÍÓÚäëïöüÄËÏÖÜÑ','aeiouAEIOUa
eiouAEIOUN'));
- comp_sede(2014,2015,2016) a la tabla comp_sede2014_2016.
INSERT INTO comp_sede2014_20166
(id,codigo_dane_estab,codigo_dane_sede,jornada,nombre)
select * from comp_sede2014 where
id not in (select id from comp_sede2014_20166);

INSERT INTO comp_sede2014_20166
(id,codigo_dane_estab,codigo_dane_sede,jornada,nombre)
select * from comp_sede2015 where
id not in (select id from comp_sede2014_20166);
select id,count(*) from comp_sede2014_20166 group by 1 order by 1;
- Plausibles(2014,2015,2016) a la tabla plausibles_2014_2016.

```

insert into plausibles_2014_2016
(select
estu_consecutivo,grupo,n,estrato,aplicacion,grado,jornada,establecimiento,e
nteterr,
departamento,municipio,sexo,sector,zona,zonastab,calendario,nivel,modelo
edu,
disenso,leng_copietas,leng_weight,leng_score1,leng_score2,leng_score3,
leng_score4,leng_score5,mate_copietas,mate_weight,mate_score1,mate_s
core2,
mate_score3,mate_score4,mate_score5,cien_copietas,cien_weight,cien_sc
ore1,
cien_score2,cien_score3,cien_score4,cien_score5,comp_copietas,comp_w
eight,
comp_score1,comp_score2,comp_score3,comp_score4,comp_score5,ano
from plausibles2014);

```

```

insert into plausibles_2014_2016
(select
estu_consecutivo,grupo,n,estrato,aplicacion,grado,jornada,establecimiento,e
nteterr,
departamento,municipio,sexo,sector,zona,zonastab,calendario,nivel,modelo
edu,
disenso,leng_copietas,leng_weight,leng_score1,leng_score2,leng_score3,
leng_score4,leng_score5,mate_copietas,mate_weight,mate_score1,mate_s
core2,
mate_score3,mate_score4,mate_score5,cien_copietas,cien_weight,cien_sc
ore1,
cien_score2,cien_score3,cien_score4,cien_score5,comp_copietas,comp_w
eight,
comp_score1,comp_score2,comp_score3,comp_score4,comp_score5,ano
from plausibles2015);

```

10) Eliminar atributos leng_score(1,2,3,4,5), mate_score(1,2,3,4,5), cien_score(1,2,3,4,5), comp_score(1,2,3,4,5). De la tabla plausibles_2014_2016

```

alter table plausibles_2014_2016 drop leng_score1, leng_score2,
leng_score3, leng_score4, leng_score5, mate_score1, mate_score2,

```

mate_score3, mate_score4, mate_score5, cien_score1, cien_score2,
cien_score3, cien_score4, cien_score5, comp_score1, comp_score2,
comp_score3, comp_score4, comp_score5;

ANEXO 2. LIMPIEZA Y TRANSFORMACIÓN DE DATOS

1. En la tabla comp_ent_terr2014_2016

- Estandarizar los valores de los atributos munexclu, nombre de minúsculas a mayúsculas y sin tildes.
update comp_ent_2014_2016 set
nombre=upper(translate(nombre,'áéíóúÁÉÍÓÚäëïöüÄËÏÖÜÑ','aeiouAEIOUaeiouAEIOUN'));

```
update comp_ent_2014_2016 set  
munexclu=upper(translate(munexclu,'áéíóúÁÉÍÓÚäëïöüÄËÏÖÜÑ','aeiouAEIOUaeiouAEIOUN'));
```

- Reemplazar los valores nulos del atributo munexclu por SIN DATO.
update comp_ent_terr2014_2016 set munexclu='SIN_DATO' where
munexclu is null

2. En la tabla comp_estab2014_2016

- Crear la tabla comp_estab2014_2016_limpio a partir de comp_estab2014_2016.
create table comp_estab2014_2016_limpio as select * from
comp_estab2014_2016;

- Estandarizar los valores del atributo nombre de minúsculas a mayúsculas y sin tildes.
update comp_estab2014_2016_limpio set
nombre=upper(translate(nombre,'áéíóúÁÉÍÓÚäëïöüÄËÏÖÜÑ','aeiouAEIOUaeiouAEIOUN'));

- Reemplazar los valores numéricos del atributo zona por su descripción de acuerdo a la siguiente tabla:

CODIGO	DESCRIPCION
1	'URBANA'
2	'RURAL'

```
update comp_estab2014_2016_limpio set zona='URBANA' WHERE  
zona='1';
```



```
update comp_estab2014_2016_limpio set zona='RURAL' WHERE zona='2';
```

- Reemplazar los valores numéricos del atributo sector por su descripción de acuerdo a la siguiente tabla:

CODIGO	DESCRIPCION
1	'OFICIAL'
2	'NO_OFICIAL'

```
update comp_estab2014_2016_limpio set sector='OFICIAL' WHERE sector='1';  
update comp_estab2014_2016_limpio set sector='NO_OFICIAL' WHERE sector='2';
```

- Reemplazar los valores numéricos del atributo tipo_estab por su descripción de acuerdo a la siguiente tabla:

CODIGO	DESCRIPCION
1	'OFICIAL URBANO'
2	'OFICIAL RURAL'
3	'NO_OFICIAL'

```
update comp_estab2014_2016_limpio set tipo_estab='OFICIAL URBANO'  
WHERE tipo_estab='1';  
update comp_estab2014_2016_limpio set tipo_estab='OFICIAL RURAL'  
WHERE tipo_estab='2';  
update comp_estab2014_2016_limpio set tipo_estab='NO OFICIAL'  
WHERE tipo_estab='3';
```

3. En la tabla comp_municipio2014_2016

- Estandarizar los valores del atributo nombre de minúsculas a mayúsculas y sin tildes.

```
update comp_municipio2014_2016 set  
nombre=upper(translate(nombre,'áéíóúÁÉÍÓÚäëïöüÄËÏÖÜÑ','aeiouAEIOUa  
eiouAEIOUN'));
```

4. En la tabla plausibles2014_2016

- Crear la tabla `plausibles_2014_2016_limpio` a partir de `comp_estab2014_2016`.
`create table plausibles_2014_2016_limpio as select * from plausibles_2014_2016;`

- Reemplazar los valores numéricos del atributo `sexo` por su descripción de acuerdo a la siguiente tabla:

CODIGO	DESCRIPCIÓN
1	MASCULINO
2	FEMENINO
3	NO_ESPECIFICA

- Reemplazar los valores numéricos del atributo `zona` por su descripción de acuerdo a la siguiente tabla:

CODIGO	DESCRIPCIÓN
1	URBANA
2	RURAL

- Reemplazar los valores numéricos del atributo `dissenso` por su descripción de acuerdo a la siguiente tabla:

CODIGO	DESCRIPCIÓN
0	SIN DISCAPACIDAD
1	CON DISCAPACIDAD

- Reemplazar los valores numéricos del atributo `leng_copietas` por su descripción de acuerdo a la siguiente tabla:

CODIGO	DESCRIPCIÓN
0	NO COPIA
1	CON COPIA
2	NO ESPECIFICA

NULL	SIN DATO
------	----------

- Reemplazar los valores Null del atributo leng_weight por SIN DATO.
- Reemplazar los valores numéricos del atributo mate_copietas por su descripción de acuerdo a la siguiente tabla:

CODIGO	DESCRIPCIÓN
0	NO COPIA
1	CON COPIA
2	NO ESPECIFICA
3	NO ESPECIFICA
NULL	SIN DATO

- Reemplazar los valores Null del atributo mate_weight por SIN DATO.
- Reemplazar los valores numéricos del atributo cien_copietas por su descripción de acuerdo a la siguiente tabla:

CODIGO	DESCRIPCIÓN
0	NO COPIA
1	CON COPIA
2	NO ESPECIFICA
3	NO ESPECIFICA
NULL	SIN DATO

- Reemplazar los valores Null del atributo cien_weight por SIN DATO.
- Reemplazar los valores numéricos del atributo comp_copietas por su descripción de acuerdo a la siguiente tabla:

CODIGO	DESCRIPCIÓN
0	NO COPIA
1	CON COPIA

NULL	SIN DATO
------	----------

- Reemplazar los valores Null del atributo *comp_weight* por SIN DATO.
- Crear la tabla *plausibles_2014_2016_transformado* a partir de *comp_estab2014_2016*.

```
create table plausibles_2014_2016_transformado as select * from
plausibles_2014_2016_limpio;
```

- Estandarizar a punto decimal los valores del atributo *leng_weight*.
update plausibles_2014_2016_transformado set leng_weight =
replace(leng_weight, ',', '.') WHERE leng_weight like '%,%';
- Estandarizar a punto decimal los valores del atributo *mate_weight*.
update plausibles_2014_2016_transformado set mate_weight =
replace(mate_weight, ',', '.') WHERE mate_weight like '%,%';
- Estandarizar a punto decimal los valores del atributo *cien_weight*.
update plausibles_2014_2016_transformado set cien_weight =
replace(cien_weight, ',', '.') WHERE cien_weight like '%,%';
- Estandarizar a punto decimal los valores del atributo *comp_weight*.
update plausibles_2014_2016_transformado set comp_weight =
replace(comp_weight, ',', '.') WHERE comp_weight like '%,%';
- Actualizar los valores anómalos del atributo *establecimiento* (que no corresponden a ningún código DANE en la tabla *comp_estab2014_2016*) teniendo en cuenta el atributo *jornada* que corresponde a la sede del establecimiento en la tabla *comp_sede2014_2016* por el cual se puede verificar el correcto establecimiento al cual pertenece.

ETABLECIMIENTO (Codigo anomalo)	ETABLECIMIENTO (Codigo corregido)	JORNADA (Codigo sede)
152215000000	152215000138	459080
152215000000	152215000138	464870

152227000000	152227000222	463388
152227000000	152227000028	463466
152317000000	152317000035	481199
152323000000	152323000161	487888
152356000000	152356000182	481168
152356000000	152356000212	496319
152356000000	152356000212	500062
152356000000	152356000204	497517
152356000000	152356000409	476824
152356000000	152356000182	481167
152356000000	152356000166	495003
152356000000	152356000581	472356
152356000000	152356000191	499095
152356000000	152356000221	285298
152356000000	152356000191	463048
152356000000	152356000409	459147
152356000000	152356001901	460903
152356000000	152356001901	463870
152573000000	152573000093	474148
152585000000	152585000161	469029
152585000000	152585000374	496506
252210000000	252210000096	478108
252215000000	252215000167	466513
252215000000	252215000035	459478
252215000000	252215000469	461799

252224000000	252224000145	498999
252227000000	252227000936	482311
252227000000	252227000057	474254
252227000000	252227000791	467188
252227000000	252227000863	470659
252227000000	252227000880	478094
252317000000	252317000081	464161
252317000000	252317000412	484482
252323000000	252323000190	484787
252352000000	252352000140	485315
252352000000	252352000166	489631
252356000000	252356000535	461878
252356000000	252356001001	465744
252356000000	252356001035	470549
252356000000	252356001019	495004
252560000000	252560000030	464144
252560000000	252560000048	491792
252585000000	252585000352	496180
352287000000	352287000018	454887
352356000000	352356000149	455949
352356000000	352356000173	185901
352356000000	352356000033	492910
352356000000	352356000149	495900
352356000000	352356001421	463860
352356000000	352356000131	499884

352356000000	352356000157	490708
352356000000	352356000467	288520
452022000000	452022000118	488781
452215000000	452215000441	496167
452215000000	452215000492	454826

update plausibles_2014_2016_transformado set
etabecimiento='152215000138' where etabecimiento='152215000000' and
jornada='459080';

update plausibles_2014_2016_transformado set
etabecimiento='152215000138' where etabecimiento='152215000000' and
jornada='464870';

update plausibles_2014_2016_transformado set
etabecimiento='152227000222' where etabecimiento='152227000000' and
jornada='463388';

update plausibles_2014_2016_transformado set
etabecimiento='152227000028' where etabecimiento='152227000000' and
jornada='463466';

update plausibles_2014_2016_transformado set
etabecimiento='152317000035' where etabecimiento='152317000000' and
jornada='481199';

update plausibles_2014_2016_transformado set
etabecimiento='152323000161' where etabecimiento='152323000000' and
jornada='487888';

update plausibles_2014_2016_transformado set
etabecimiento='152356000182' where etabecimiento='152356000000' and
jornada='481168';

update plausibles_2014_2016_transformado set
establecimiento='152356000212' where establecimiento='152356000000' and
jornada='496319';

update plausibles_2014_2016_transformado set
establecimiento='152356000212' where establecimiento='152356000000' and
jornada='500062';

update plausibles_2014_2016_transformado set
establecimiento='152356000204' where establecimiento='152356000000' and
jornada='497517';

update plausibles_2014_2016_transformado set
establecimiento='152356000409' where establecimiento='152356000000' and
jornada='476824';

update plausibles_2014_2016_transformado set
establecimiento='152356000182' where establecimiento='152356000000' and
jornada='481167';

update plausibles_2014_2016_transformado set
establecimiento='152356000166' where establecimiento='152356000000' and
jornada='495003';

update plausibles_2014_2016_transformado set
establecimiento='152356000581' where establecimiento='152356000000' and
jornada='472356';

update plausibles_2014_2016_transformado set
establecimiento='152356000191' where establecimiento='152356000000' and
jornada='499095';

update plausibles_2014_2016_transformado set
establecimiento='152356000221' where establecimiento='152356000000' and
jornada='285298';

update plausibles_2014_2016_transformado set
establecimiento='152356000191' where establecimiento='152356000000' and
jornada='463048';

update plausibles_2014_2016_transformado set
establecimiento='152356000409' where establecimiento='152356000000' and
jornada='459147';

update plausibles_2014_2016_transformado set
establecimiento='152356001901' where establecimiento='152356000000' and
jornada='460903';

update plausibles_2014_2016_transformado set
establecimiento='152356001901' where establecimiento='152356000000' and
jornada='463870';

update plausibles_2014_2016_transformado set
establecimiento='152573000093' where establecimiento='152573000000' and
jornada='474148';

update plausibles_2014_2016_transformado set
establecimiento='152585000161' where establecimiento='152585000000' and
jornada='469029';

update plausibles_2014_2016_transformado set
establecimiento='152585000374' where establecimiento='152585000000' and
jornada='496506';

update plausibles_2014_2016_transformado set
establecimiento='252210000096' where establecimiento='252210000000' and
jornada='478108';

update plausibles_2014_2016_transformado set
establecimiento='252215000167' where establecimiento='252215000000' and
jornada='466513';

update plausibles_2014_2016_transformado set
establecimiento='252215000035' where establecimiento='252215000000' and
jornada='459478';

update plausibles_2014_2016_transformado set
establecimiento='252215000469' where establecimiento='252215000000' and
jornada='461799';

update plausibles_2014_2016_transformado set
establecimiento='252224000145' where establecimiento='252224000000' and
jornada='498999';

update plausibles_2014_2016_transformado set
establecimiento='252227000936' where establecimiento='252227000000' and
jornada='482311';

update plausibles_2014_2016_transformado set
establecimiento='252227000057' where establecimiento='252227000000' and
jornada='474254';

update plausibles_2014_2016_transformado set
establecimiento='252227000791' where establecimiento='252227000000' and
jornada='467188';

update plausibles_2014_2016_transformado set
establecimiento='252227000863' where establecimiento='252227000000' and
jornada='470659';

update plausibles_2014_2016_transformado set
establecimiento='252227000880' where establecimiento='252227000000' and
jornada='478094';

update plausibles_subregion_obando set establecimiento='252317000081'
where establecimiento='252317000000' and jornada='464161';

update plausibles_2014_2016_transformado set
establecimiento='252317000412' where establecimiento='252317000000' and
jornada='484482';

update plausibles_2014_2016_transformado set
establecimiento='252323000190' where establecimiento='252323000000' and
jornada='484787';

update plausibles_2014_2016_transformado set
establecimiento='252352000140' where establecimiento='252352000000' and
jornada='485315';

update plausibles_2014_2016_transformado set
establecimiento='252352000166' where establecimiento='252352000000' and
jornada='489631';

update plausibles_2014_2016_transformado set
establecimiento='252356000535' where establecimiento='252356000000' and
jornada='461878';

update plausibles_2014_2016_transformado set
establecimiento='252356001001' where establecimiento='252356000000' and
jornada='465744';

update plausibles_2014_2016_transformado set
establecimiento='252356001035' where establecimiento='252356000000' and
jornada='470549';

update plausibles_2014_2016_transformado set
establecimiento='252356001019' where establecimiento='252356000000' and
jornada='495004';

update plausibles_2014_2016_transformado set
establecimiento='252560000030' where establecimiento='252560000000' and
jornada='464144';

update plausibles_2014_2016_transformado set
establecimiento='252560000048' where establecimiento='252560000000' and
jornada='491792';

update plausibles_2014_2016_transformado set
establecimiento='252585000352' where establecimiento='252585000000' and
jornada='496180';

update plausibles_2014_2016_transformado set
establecimiento='352287000018' where establecimiento='352287000000' and
jornada='454887';

update plausibles_2014_2016_transformado set
establecimiento='352356000149' where establecimiento='352356000000' and
jornada='455949';

update plausibles_2014_2016_transformado set
establecimiento='352356000173' where establecimiento='352356000000' and
jornada='185901';

update plausibles_2014_2016_transformado set
establecimiento='352356000033' where establecimiento='352356000000' and
jornada='492910';

update plausibles_2014_2016_transformado set
establecimiento='352356000149' where establecimiento='352356000000' and
jornada='495900';

update plausibles_2014_2016_transformado set
establecimiento='352356001421' where establecimiento='352356000000' and
jornada='463860';

update plausibles_2014_2016_transformado set
establecimiento='352356000131' where establecimiento='352356000000' and
jornada='499884';

update plausibles_2014_2016_transformado set
establecimiento='352356000157' where establecimiento='352356000000' and
jornada='490708';

update plausibles_2014_2016_transformado set
establecimiento='352356000467' where establecimiento='352356000000' and
jornada='288520';

update plausibles_2014_2016_transformado set
establecimiento='452022000118' where establecimiento='452022000000' and
jornada='488781';

```
update plausibles_2014_2016_transformado set
establecimiento='452215000441' where establecimiento='452215000000' and
jornada='496167';
```

```
update plausibles_2014_2016_transformado set
establecimiento='452215000492' where establecimiento='452215000000' and
jornada='454826';
```

- Actualizar los valores anómalos del atributo *jornada* (que no corresponden a ningún código DANE en la tabla *comp_sede2014_2016*) teniendo en cuenta el atributo *establecimiento* que corresponde a la institución en la tabla *comp_estab2014_2016* por el cual se puede verificar la correcta jornada a la cual pertenece.

JORNADA (Codigo anomalo)	JORNADA (Codigo corregido)	ETABLECIMIENTO (Codigo DANE)
462193	542180	252227000162
465335	627537	252356000101

```
update plausibles_2014_2016_transformado set jornada='542180' where
establecimiento='252227000162' and jornada='462193';
```

```
update plausibles_2014_2016_transformado set jornada='627537' where
establecimiento='252356000101' and jornada='465335';
```